CANADIAN JOURNAL OF MATHEMATICS

Journal Canadien de Mathématiques

VOL. XII - NO. 3
1960

Published for
THE CANADIAN MATHEMATICAL CONGRESS
by the
University of Toronto Press

The chief languages of the *Journal* are English and French.

Manuscripts for publication in the *Journal* should be sent to the *Editor-in-Chief*, G. F. D. Duff, University of Toronto. Authors are asked to write with a sense of perspective and as clearly as possible, especially in the introduction. Regarding typographical conventions, attention is drawn to the Author's Manual of which a copy will be furnished on request.

All other correspondence should be addressed to the *Managing Editor*, G. de B. Robinson, University of Toronto.

The *Journal* is published quarterly. Subscriptions should be sent to the *Managing Editor*. The price per volume of four numbers is $10.00. This is reduced to $5.00 for individual members of recognized Mathematical Societies.

The Canadian Mathematical Congress gratefully acknowledges the assistance of the following towards the cost of publishing this Journal:

University of Alberta	Assumption University
University of British Columbia	Carleton College
Dalhousie University	Ecole Polytechnique
Université Laval	Loyola College
University of Manitoba	McGill University
McMaster University	Université de Montréal
Mount Allison University	Nova Scotia Technical College
Queen's University	St. Mary's University
University of Saskatchewan	University of Toronto

National Research Council of Canada
and the
American Mathematical Society

AUTHORIZED AS SECOND CLASS MAIL, POST OFFICE DEPARTMENT, OTTAWA

A CLASS OF FUNCTION ALGEBRAS

F. W. ANDERSON

Introduction. A problem which has generated considerable interest during the past couple of decades is that of characterizing abstractly systems of real-valued continuous functions with various algebraic or topological-algebraic structures. With few exceptions known characterizations are of systems of bounded continuous functions on compact or locally compact spaces. Only recently have characterizations been given of the systems $C(X)$ of all real-valued continuous functions on an arbitrary completely regular space X **(1)**. One of the main objects of this paper is to provide, by using certain special techniques, a characterization of $C(X)$ for a particular class of (not necessarily compact) completely regular spaces.

Generally speaking, one of the primary difficulties in characterizing all of $C(X)$ is that of obtaining conditions which insure that a subsystem is, in fact, all of $C(X)$. Sets of conditions of two different types have evolved. The first, for X compact, uses the completeness of $C(X)$ in its usual norm and the Stone-Weierstrass Theorem. (For example, see **(10)** and **(13)**.) The second uses the fact that $C(X)$ is, in a sense, maximal in a certain class of algebraic systems (cf. **(1, 6)**. The first of these appears to be applicable only in situations where $C(X)$ possesses a norm or a suitable family of pseudo-norms. The second, although it applies in more general situations and is algebraic in nature, has the slight drawback of the "external" character of the maximality condition.

In this paper we characterize $C(X)$ as a vector lattice, as an l-ring, and as an algebra[1] for the case in which X is a P-space **(7)**. A feature of special interest in these characterizations is that we appeal to neither of the aforementioned methods for obtaining all of $C(X)$; rather we use, for X a P-space, a simple property of certain "fixed" subsets of $C(X)$. En route to obtaining these results we also characterize $M(X, \mathfrak{B})$, the set of all real-valued measurable functions on a total measurable space, as a vector lattice and as an l-ring.

In two recent papers, Brainerd (**(4)** and **(5)**) has also given characterizations of $C(X)$, X a P-space, and $M(X, \mathfrak{B})$ as l-algebras. The characterizations of $C(X)$ by Brainerd as an l-algebra and by us as an l-ring, although obtained independently, use essentially the same techniques.

Received April 5, 1958. Presented to the American Mathematical Society June 20, 1958.

[1]For the theory of vector lattices and l-rings see Birkhoff **(2)**, Birkhoff and Pierce **(3)**, and Nakano **(11)**. Our notation will be that of **(2)** except that $\vee$ and $\wedge$ will be used to denote lattice join and meet, respectively. By an algebra we shall always mean an algebra over the real field.

1. Preliminaries. If X is a set, denote by $F(X)$ the set of all real-valued functions on X. If $\mathfrak{B}$ is a Boolean σ-algebra of subsets of X, then we say that the pair $(X, \mathfrak{B})$ is a *total measurable space* and denote by $M(X, \mathfrak{B})$ the set of all $f \in F(X)$ measurable $\mathfrak{B}$. If $\mathfrak{T}$ is a base for a topology on X, then we denote by $C(X, \mathfrak{T})$, or in unambiguous cases simply $C(X)$, the set of all $f \in F(X)$ continuous with respect to $\mathfrak{T}$.

For each $f \in F(X)$ set $Z(f) = \{x \in X; f(x) = 0\}$. A subset $Z \subseteq X$ is a measurable zero set in case $Z \in \mathfrak{B}$, or equivalently, in case $Z = Z(f)$ for some $f \in M(X, \mathfrak{B})$. A subset $Z \subseteq X$ is a *continuous zero set* in case $Z = Z(f)$ for some $f \in C(X)$.

A subset $I \subseteq F(X)$ is *fixed* in case $\bigcap \{Z(f); f \in I\}$, also written $\bigcap Z(I)$, is non-empty. Let $A \subseteq F(X)$. A set $I \subseteq A$ is a maximal fixed subset of A if and only if $I = \{f \in A; f(x) = 0\}$ for some $x \in X$. In general, different points in X do not give rise to different maximal fixed subsets of A; if, however, A separates points (that is, $x \neq y$ in X implies $0 = f(x) \neq f(y)$ for some $f \in A$), then the mapping $I \to \bigcap Z(I)$ is one-one from the maximal fixed subsets of A onto X.

Let $A \subseteq F(X)$. Then for each $I \subseteq A$, set

$$\mathfrak{a}(I) = \{f \in A; X - \bigcap Z(I) \subseteq Z(f)\}.$$

Thus $f \in \mathfrak{a}(I)$ if and only if for every $x \in X$ and every $g \in I$, $f(x)g(x) = 0$. We say that $I \subseteq A$ is *Z-convex* (in A) provided that

$$I = \{f \in A; \bigcap Z(I) \subseteq Z(f)\}.$$

It is clear then that for each $I \subseteq A$, if $I = \mathfrak{a}(\mathfrak{a}(I))$, then I is Z-convex. The converse in general is false; for example, every maximal fixed subset I of A is Z-convex, but it need not satisfy $I = \mathfrak{a}(\mathfrak{a}(I))$. If $\mathscr{S}$ is the collection ofmaximal fixed subsets of A, then it is clear that $I \subseteq A$ is Z-convex if and only if $I = \bigcap \{N \in \mathscr{S}; I \subseteq N\}$.

A topological space X $(= (X, \mathfrak{T}))$ is a *P-space* **(7)** provided that X is completely regular and that every G_δ-set in X is open. In such a space X the family of continuous zero sets of X is an open base for the topology, is a Boolean σ-algebra of subsets of X, and coincides with the family of closed-open subsets of X. Conversely, if $(X, \mathfrak{B})$ is a total measurable space which separates points of X (that is, $x \neq y$ in X implies $x \in E$ and $y \notin E$ for some $E \in \mathfrak{B}$), then $\mathfrak{B}$ is an open base for a topology on X relative to which X is a P-space. Moreover, it is clear that in this case $M(X, \mathfrak{B}) \subseteq C(X, \mathfrak{B})$. We now prove a test for equality.[2]

LEMMA 1.1. *Let X be a P-space, let $\mathfrak{B}$, a Boolean σ-algebra of subsets of X, be an open base for the topology of X, and let $\mathscr{S}$ be the set of maximal fixed subsets of $M(X, \mathfrak{B})$. Then $M(X, \mathfrak{B}) = C(X, \mathfrak{B})$ if and only if for every Z-convex set $I \subseteq M(X, \mathfrak{B})$, if $N \in \mathscr{S}$ implies $I \not\subseteq N$ or $\mathfrak{a}(I) \not\subseteq N$, then $I = \mathfrak{a}(f)$ for some $f \in M(X, \mathfrak{B})$.*

[2] See also **(5, Theorem 1)** for a variation of this result.

Proof. Let $I \subseteq M(X, \mathfrak{B})$ be Z-zoncvx. We shall prove first that the two conditions

(1) $I \not\subset N$ or $\mathfrak{a}(I) \not\subset N$ for all $N \in \mathscr{S}$;

(2) $\cap Z(I)$ is a continuous zero set;

are equivalent. Assume (1). Then $F_1 = \cap Z(I)$ and $F_2 = \cap Z(\mathfrak{a}(I))$ are disjoint. For let $x \in X$ and let

$$N_x = \{f \in M(X, \mathfrak{B}); f(x) = 0\}.$$

Since N_x is Z-convex, we have $x \in F_1$ if and only if $I \subseteq N_x$, and $x \in F_2$ if and only if $\mathfrak{a}(I) \subseteq N_x$. Thus, by (1), $F_1 \cap F_2 = \phi$. Since $M(X, \mathfrak{B}) \subseteq C(X, \mathfrak{B})$, we conclude that F_1 and F_2 are closed. Since $\mathfrak{B}$ is an open base, if $x \in X$, then $\{x\} = \cap Z(N_x)$. Therefore if $x \notin F_1$ and if $f \in M(X, \mathfrak{B})$, then $\{x\} = \cap Z(N_x) \subseteq X - F_1 \subseteq Z(f)$ implies $f \in N_x$. That is, $\mathfrak{a}(I) \subseteq N_x$, so that $x \in F_2$. Hence $X = F_1 \cup F_2$. We have then that F_1 is both closed and open, and therefore $F_1 = \cap Z(I)$ is a continuous zero set. Conversely, assume (2). Then since X is a P-space, $F = \cap Z(I)$ is closed and open. Since $\mathfrak{B}$ is an open base, if $x \in F$, then there is an $f \in M(X, \mathfrak{B})$ such that $f(x) \neq 0$ and $X - F \subseteq Z(f)$; that is, $f \in \mathfrak{a}(I)$ and $f \notin N_x$. Hence $I \subseteq N_x$ implies $\mathfrak{a}(I) \not\subset N_x$. Thus (1) and (2) are equivalent.

We now easily prove the "only if" portion of the lemma. For suppose that $M(X, \mathfrak{B}) = C(X, \mathfrak{B})$ and that $I \subseteq M(X, \mathfrak{B})$ satisfies (1). Then $F = \cap Z(I)$ is closed and open so that the characteristic function f of F is in $M(X, \mathfrak{B})$. It is evident then that $I = \mathfrak{a}(1 - f)$.

Conversely, let $g \in C(X, \mathfrak{B})$ and let α be a real number. Set $Z = \{x \in X; g(x) \geqq \alpha\}$. Then $Z = Z((\alpha - g) \vee 0)$ is a continuous zero set. Let

$$I = \{f \in M(X, \mathfrak{B}); Z \subseteq Z(f)\}.$$

Then I is Z-convex and $\cap Z(I) = Z$. Therefore I satisfies (2) and hence (1). Thus, if $M(X, \mathfrak{B})$ satisfies the condition of the lemma, $I = \mathfrak{a}(f)$ for some $f \in M(X, \mathfrak{B})$. We claim that $Z = X - Z(f)$. Certainly $X - Z(f) \subseteq Z$. Suppose then that $x \in Z(f)$. Since $f \in M(X, \mathfrak{B}) \subseteq C(X, \mathfrak{B})$, $Z(f)$ is a continuous zero set, and therefore, since X is a P-space, $Z(f)$ is open. Now $\mathfrak{B}$ is an open base, so there is an $h \in M(X, \mathfrak{B})$ such that $h(x) \neq 0$ and $X - Z(f) \subseteq Z(h)$. Then $h \in I$ and $Z = \cap Z(I) \subseteq Z(h)$. Hence $x \notin Z$, and we have the desired reverse inclusion $X - Z(f) \supseteq Z$. Now $Z(f)$ is measurable since $f \in M(X, \mathfrak{B})$, and therefore its complement Z is measurable. Consequently, since α was arbitrary, we conclude that g is measurable $\mathfrak{B}$ and hence that $g \in M(X, \mathfrak{B})$. Thus $M(X, \mathfrak{B}) = C(X, \mathfrak{B})$ and the lemma is proved.

2. Vector lattices of functions. In this section we characterize $M(X, \mathfrak{B})$ and $C(X, \mathfrak{T})$ abstractly as vector lattices where $(X, \mathfrak{B})$ is a total measurable space and $(X, \mathfrak{T})$ is a P-space.

Let A be a vector lattice. For $f, g \in A$ we write $f \perp g$ in case $|f| \wedge |g| = 0$. A countable set $\{f_n\}$ of elements of A is a *$\sigma^\perp$-set* in case $f_n \geqq 0$ $(n = 1, 2 \ldots)$

and for each $n \neq m$, $f_n \perp f_m$. We say that A is $\sigma^\perp$-*complete* in case every $\sigma^\perp$-set $\{f_n\}$ in A has a least upper bound, $\vee_n f_n$, in A.[3]

LEMMA 2.1. *Let A be a vector sublattice of $F(X)$ which separates points of X and contains the constant function 1. Then $A = M(X, \mathfrak{B})$ for some point separating σ-algebra $\mathfrak{B}$ of subsets of X if and only if A is $\sigma^\perp$-complete and σ-complete.*

Proof. The necessity of these conditions follows readily from the fact that if $A = M(X, \mathfrak{B})$, then the desired countable spurema are simply the "pointwise" suprema.

Conversely, let A satisfy the stated conditions. If $\{f_n\} \subseteq A$ with $f = \vee_n f_n \in A$, then we claim that $f(x) = \vee_n[f_n(x)]$ for each $x \in X$. For suppose, on the contrary, that there is an $x \in X$ with $f(x) > \vee_n[f_n(x)]$. Without loss of generality, we may assume that, for all n, $0 < f_n \leqq f_{n+1} < 1$ and $f_n(x) = 0$, and that $f(x) = 1$. Now define sequences $\{g_n\}$, $\{h_n\}$, and $\{e_n\}$ in A by

$$g_1 = 2f_1 \wedge 1 \quad \text{and} \quad g_n = 2(f_n \vee g_{n-1}) \wedge 1 \quad \text{for} \quad n > 1;$$
$$h_n = (2g_n - g_{n+1})^+;$$

and

$$e_1 = h_1, \quad e_2 = 2h_2, \quad \text{and} \quad e_n = n(h_n - h_{n-2}) \quad \text{for} \quad n > 2.$$

Also, for each n, set

$$Y_n = \{y \in X; g_n(y) = 1\}.$$

Then one easily shows that, for each n, $0 \leqq h_n \leqq h_{n+1} \leqq 1$, $h_n(Y_n) = 1$, and $h_n(X - Y_{n+1}) = 0$. From these it follows that $0 \leqq e_n \leqq n$, $e_n(x) = 0$, $e_n(Y_n - Y_{n-1}) = n$, and

$$X - Z(e_n) \subseteq Y_{n+1} - Y_{n-2},$$

where $Y_{-1} = Y_0 = \phi$. This implies that if $|m - n| > 2$, then $e_m \perp e_n$; hence each of the sets $\{e_{3n}\}$, $\{e_{3n-1}\}$, and $\{e_{3n-2}\}$ is a $\sigma^\perp$-set in A. Therefore, since A is $\sigma^\perp$-complete,

$$e = \bigvee_{i=0}^{2} \left(\bigvee_{n=1}^{\infty} e_{3n-i} \right) = \bigvee_{n=1}^{\infty} e_n$$

is in A. Now if $f_n(y) > 0$, then $2^k f_n(y) \geqq 1$ for some k; therefore, since

$$g_{n+k}(y) \geqq [2^k f_n(y)] \wedge 1 = 1,$$

we have $y \in Y_{n+k}$. That is, if

$$P = \bigcup_{n=1}^{\infty} (X - Z(f_n)),$$

then

$$P \subseteq \bigcup_{n=1}^{\infty} Y_n = \bigcup_{n=0}^{\infty} (Y_n - Y_{n-1}).$$

[3]Other, possibly less descriptive, terminology for this notion includes σ-full **(2)** and complete **(11)**.

Thus we have that $e \geqq 1$ on P, and consequently, that $f \leqq e$ on X. Hence there is an integer $k \geqq 2$ such that $e(x) \leqq k - 1$. Set

$$e' = \left(\bigvee_{i=1}^{k-1} ke_i \right) \vee e.$$

Then $e'(y) \geqq k$ for all $y \in P$ and $e'(x) \leqq e(x) \leqq k$. Therefore $(e' - k + 1)^+ \geqq 1$ on P and, as a result, $f \leqq (e' - k + 1)^+$. This is a contradiction since $(e' - k + 1)^+(x) = 0$. We conclude then that $f(x) = 0$, and therefore countable suprema in A, when defined, are defined pointwise.

For each $f \in A$, set $e_f = \vee_n (|nf| \wedge 1)$; then, by the result of the preceding paragraph, e_f is the characteristic function of $X - Z(f)$. Thus A contains e_f and $1 - e_f$ the characteristic functions of $X - Z(f)$ and $Z(f)$, respectively. Now let $\mathfrak{B} = \{Z(f); f \in A\}$. Then $\mathfrak{B}$ is an algebra of subsets of X; for $Z(f) \cup Z(g) = Z(|f| \wedge |g|)$ and $X - Z(f) = Z(1 - e_f)$. Since A is point separating, it is clear that $\mathfrak{B}$ also is point separating. Moreover, $\mathfrak{B}$ is a σ-algebra; for, using the result of the first paragraph and the σ-completeness of A, we have

$$\cap_n Z(f_n) = \cap_n Z(e_{f_n}) = Z(\vee_n e_{f_n}) \in \mathfrak{B}.$$

We show next that $A \subseteq M(X, \mathfrak{B})$. Let $f \in A$ and let α be real. Then

$$\{x \in X; f(x) \geqq \alpha\} = Z((\alpha - f)^+) \in \mathfrak{B},$$

so that $f \in M(X, \mathfrak{B})$. On the other hand, A contains all measurable characteristic functions, and so, since A is σ-complete, A contains all bounded $f \in M(X, \mathfrak{B})$. (Cf. **(8**, Theorem 20.B**)**.) To complete the proof we need only show that A contains all non-negative $f \in M(X, \mathfrak{B})$. So let $f \geqq 0$ in $M(X, \mathfrak{B})$. For each $n = 1, 2, \ldots$, set

$$E_n = \{x \in X; n - 1 \leqq f(x) < n\}$$

and let $f_n \in F(X)$ be defined by $f_n = f$ on E_n and $f_n = 0$ on $X - E_n$. Then obviously $f_n \in M(X, \mathfrak{B})$ and is bounded; hence $f_n \in A$ for all n. But $\{f_n\}$ is a $\sigma^\perp$-set, so that $f = \vee_n f_n \in A$. Thus the proof of the lemma is complete.

It is interesting to note that neither $\sigma^\perp$-completeness nor σ-completeness alone is adequate to insure that $A = M(X, \mathfrak{B})$. For example, if X is uncountable, then the set of all $f \in F(X)$ with $f(X)$ countable is a vector sublattice of $F(X)$ which is $\sigma^\perp$-complete but not σ-complete. Next let X be the Stone-Čech compactification of an infinite discrete space and let $A = C(X)$. Then A is a vector sublattice of $F(X)$ which is σ-complete but not $\sigma^\perp$-complete; in fact, there exist bounded sequences $\{f_n\}$ in A such that $Z(\vee_n f_n) \neq \cap_n Z(f_n)$.

Let A be a vector lattice. An element $e \in A$ is a *weak order unit* in case for all $f \in A$, $|f| \wedge |e| = 0$ implies $f = 0$. A subset $I \subseteq A$ is an *ideal* of A in case I is a linear subspace such that $f \in I$ and $|g| \leqq |f|$ implies that $g \in I$.

THEOREM 2.2. *A vector lattice A is isomorphic to the vector lattice $M(X, \mathfrak{B})$ for some total measurable space $(X, \mathfrak{B})$ if and only if A is $\sigma^\perp$-complete, σ-complete,*

has a weak order unit, and $\cap \mathscr{S} = 0$ *where* $\mathscr{S}$ *is the set of maximal ideals of* A. *In fact, when* A *satisfies the stated conditions,* A *is isomorphic to* $M(\mathscr{S}, \mathfrak{B})$, *where* $\mathfrak{B}$ *is a point separating* σ*-algebra of subsets of* $\mathscr{S}$.

Proof. Since the family $\mathscr{F}$ of fixed maximal ideals (= maximal fixed ideals) of $M(X, \mathfrak{B})$ satisfies $\cap \mathscr{F} = 0$, the necessity of the conditions is obvious.

Conversely, let A satisfy the stated conditions. Let $e \in A$ be a weak order unit for A; we may assume that $e \geqq 0$. We claim that if $\mathscr{T} = \{N \in \mathscr{S}; e \notin N\}$, then $\cap \mathscr{T} = 0$. For if $f \in \cap \mathscr{T}$, then, for every $N \in \mathscr{S}$, either $f \in N$ or $e \in N$. Thus $(|f| \wedge e) \in \cap \mathscr{S}$ so that $|f| \wedge e = 0$. Since e is a weak order unit, this implies $f = 0$. That is, $\cap \mathscr{T} = 0$. By a familiar technique **(1)** we can define an isomorphism of A onto a point-separating vector sublattice A^* of $F(\mathscr{T})$ such that e is mapped onto the constant function 1. Appealing to Lemma 2.1 we have that $A^* = M(\mathscr{T}, \mathfrak{B})$ for some σ-algebra $\mathfrak{B}$ of subsets of $\mathscr{T}$.

To complete the proof it will suffice to show that $\mathscr{S} = \mathscr{T}$, and for this it will suffice to show that if $(X, \mathfrak{B})$ is a total measurable space, then no maximal ideal of $M(X, \mathfrak{B})$ contains 1. Suppose, on the contrary, that N is a maximal ideal of $M(X, \mathfrak{B})$ and that $1 \in N$. Then since N is proper, there is an $f \geqq 0$ with $f \notin N$. Since N is maximal and since $f^2 > f$ is in $M(X, \mathfrak{B})$, there is a real number α such that $f^2 - \alpha f \in N$. Let $\beta = \frac{1}{4}(\alpha + 1)^2$. Since $1 \in N$, it follows that β, and hence $f^2 - \alpha f + \beta$, belongs to N. But

$$f^2 - \alpha f + \beta = [f - \tfrac{1}{2}(\alpha + 1)]^2 + f \geqq f,$$

contrary to $f \notin N$. Thus the assumption $1 \in N$ is untenable and the proof is complete.

Let A be a vector lattice and let $I \subseteq A$. We set

$$I^\perp = \{f \in A; f \perp g \text{ for all } g \in I\}.$$

Then clearly, $I \subseteq I^{\perp\perp}$. If $\mathscr{S}$ is a family of ideals of A, then an ideal I of A is *$\mathscr{S}$-complemented* in case $I = I^{\perp\perp}$, and for each $N \in \mathscr{S}$, either $I \not\subseteq N$ or $I^\perp \not\subseteq N$.

THEOREM 2.3. *Let* A *be a vector lattice and let* $\mathscr{S}$ *be the set of all maximal ideals of* A. *Then* A *is isomorphic to the vector lattice* $C(X)$ *for some completely regular* P*-space* X *if and only if* A *is* $\sigma^\perp$*-complete,* σ*-complete,* $\cap \mathscr{S} = 0$, *and for each* $\mathscr{S}$*-complemented ideal* I *of* A, $I = \{f\}^\perp$ *for some* $f \in A$.

Proof. To prove the necessity we may assume that X is a Q-space **(9)**; for if X is a P-space, then so is υX, and, of course, $C(X)$ and $C(\upsilon X)$ are isomorphic. With this assumption the maximal ideals of $C(X)$ coincide with the maximal fixed subsets of $C(X)$. Moreover, if $I \subseteq C(X)$, then $I^\perp$ coincides with the set $\mathfrak{a}(I)$ defined in § 1. These observations combine with Lemma 1.1 and Theorem 2.2 to establish the necessity of the conditions in the present theorem.

Conversely, let A satisfy the stated conditions. Since the zero ideal of A is clearly $\mathscr{S}$-complemented, it follows that A has a weak order unit. Therefore, by Theorem 2.2, A is isomorphic to $M(X, \mathfrak{B})$ for some total measurable space $(X, \mathfrak{B})$ where, in fact, the maximal ideals $\mathscr{S}$ correspond to the maximal fixed ideals of $M(X, \mathfrak{B})$. A Z-convex set I^* of $M(X, \mathfrak{B})$ is then the image of some $I = \cap \{N \in \mathscr{S}; I \subseteq N\}$ in A, and therefore is an ideal of $M(X, \mathfrak{B})$. Since we clearly have $\mathfrak{a}(I^*) = (I^*)^\perp$, it follows from Lemma 1.1 that $M(X, \mathfrak{B}) = C(X, \mathfrak{B})$, and the proof is complete.

3. f-rings of functions. In this section we characterize $M(X, \mathfrak{B})$ and $C(X, \mathfrak{T})$, $(X, \mathfrak{B})$ and $(X, \mathfrak{T})$ as before, as f-rings. Although these characterizations still require σ-completeness, we are able to dispense with the full force of the $\sigma^\perp$-completeness requirement. In its place we use ring regularity and a condition of countbable character on certain ideals. These characterizations are slightly sharpened versions of those given in **(4** and **5)**.

Recall that an f-ring **(3)** is a lattice-ordered ring A with the property that for all $f, g, h \in A$, $f \wedge g = 0$ and $h \geqq 0$ together imply $hf \wedge g = fh \wedge g = 0$. Clearly $M(X, \mathfrak{B})$ and $C(X, \mathfrak{T})$ are f-rings.

A ring A is *regular* **(12)** in case for each $f \in A$, there is an $f' \in A$ such that $ff'f = f$. It is known **(7)** that a completely regular space X is a P-space if and only if $C(X)$, as a ring, is regular. Concerning regular f-rings we prove the following result which may be of independent interest.

LEMMA 3.1. *Let A be a regular f-ring. Then*

(1) *For all $f, g \in A$, $|f| \wedge |g| = 0$ if and only if $fg = 0$.*

(2) *If A has a weak order unit, A has an identity.*

Proof. Since A has no non-zero nilpotent elements, the l-radical of A is zero **(3)**. Therefore (1) follows from **(3**, Corollary 1, p. 57**)** and **(3**, Corollary 2, p. 63**)**. Next let $e \geqq 0$ be a weak order unit for A and let $ee'e = e$. Then, by (1), $f \wedge ee' = 0$ implies $fee'e = fe = 0$ which implies $|f| \wedge e = 0$ and thus $f = 0$. That is, the idempotent $e'' = ee'$ is also a weak order unit. Let $f \in A$; then $(fe'' - f)e'' = 0$ implies $|fe'' - f| \wedge e'' = 0$. Therefore, since e'' is a weak order unit, $fe'' = f$. Similarly, $e''f = f$, which establishes (2).

An ideal I of a ring A is σ-closed in case for every countable set $\{f_n\} \subseteq I$ there is an $f \in A$ with $ff_n = f_nf = f_n$ for all n.

THEOREM 3.2. *Let A be an f-ring and let $\mathscr{S}$ be the set of σ-closed maximal ring ideals of A. Then A is isomorphic to the f-ring $M(X, \mathfrak{B})$ for some total measurable space $(X, \mathfrak{B})$ if and only if A is regular, σ-complete, has a weak order unit, and $\cap \mathscr{S} = 0$. Moreover, if A satisfies these conditions, the space $(X, \mathfrak{B})$ and the isomorphism of A onto $M(X, \mathfrak{B})$ may be so chosen that the set $\mathscr{S}$ is mapped one-one onto the maximal fixed subsets of $M(X, \mathfrak{B})$.*

Proof. The necessity of the conditions is easily proved; we omit the details. Conversely, let A satisfy the stated conditions. Then, by Lemma 3.1, A has

a ring identity e. Moreover, since A is σ-complete, it is Archimedean **(2**, p. 229**)**, and therefore A is commutative **(3**, Theorem 13**)**. Since the regular σ-complete subring of A generated by e is isomorphic to the ordered field R of real numbers, we may regard A as a regular f-algebra over R (that is, A is a regular F-ring in the sense of **(4)**). Now let $N \in \mathscr{S}$ and $\{f_n\} \subseteq N$ such that $\vee_n f_n \in A$. Since N is σ-closed, there is an $f \in A$ with $ff_n = f_n$ for all n. By the regularity of A we may assume that f is idempotent. Then **(11**, Theorem 25.1**)**, $\vee_n f_n = \vee_n ff_n = f(\vee_n f_n) \in N$. Therefore A satisfies the conditions required in Brainerd's characterization **(4**, p. 682**)**. Thus there exist a total measurable space $(X, \mathfrak{B})$ and an isomorphism of A onto $M(X, \mathfrak{B})$ with the desired properties.

Let A be a ring. For $I \subseteq A$, set $\mathfrak{a}(I) = \{f \in A; fg = 0 \text{ for all } g \in I\}$. In general $\mathfrak{a}(I)$ is a left ideal of A; if A is commutative or if A is a regular f-ring, then $\mathfrak{a}(I)$ is a two-sided ideal.[4] A left ideal I of A is *$\mathfrak{a}$-principal* in case $I = \mathfrak{a}(f)$ for some $f \in A$.

If $\mathscr{S}$ is a family of ideals of a ring A, then an ideal I of A is *$\mathscr{S}$-complemented* in case $I = \mathfrak{a}(\mathfrak{a}(I))$ and for each $N \in \mathscr{S}$, either $I \not\subset N$ or $\mathfrak{a}(I) \not\subset N$.

THEOREM 3.3. *Let A be an f-ring and let $\mathscr{S}$ be the set of σ-closed maximal ring ideals of A. Then A is isomorphic to the f-ring $C(X)$ for some P-space X if and only if A is regular, σ-complete, $\cap \mathscr{S} = 0$, and every $\mathscr{S}$-complemented ideal of A is $\mathfrak{a}$-principal.*

Proof. To prove the necessity, we may assume that X is a Q-space. Then an application of Theorem 3.2 and Lemma 1.1 completes this portion of the proof.

Conversely, since the zero ideal of A is $\mathscr{S}$-complemented, it is $\mathfrak{a}$-principal. But from $\{0\} = \mathfrak{a}(f)$ and Lemma 3.1 we conclude that $|f|$ is a weak order unit for A. Therefore, by Theorem 3.2 and Lemma 1.1, we have that A is isomorphic to $M(X, \mathfrak{B})$ and that $M(X, \mathfrak{B}) = C(X, \mathfrak{B})$ where $(X, \mathfrak{B})$ is a P-space.

4. The algebra $C(X)$. With no assumptions concerning order properties it seems to be difficult to obtain a reasonably simple characterization of the algebras $M(X, \mathfrak{B})$. It is possible, however, to characterize the algebra $C(X)$, X a P-space, and it is the object of this section to present such a characterization.

Let A be a ring and let $\mathscr{S}$ be a family of ideals of A. A set $\{f_\alpha\} \subseteq A$ is a *discrete $\mathscr{S}$-cover* in case $\alpha \neq \beta$ implies $f_\alpha f_\beta = 0$ and the set $\{f_\alpha\}$ is contained in no member of $\mathscr{S}$. We say that A is *$\mathscr{S}$-regular* in case for each discrete $\mathscr{S}$-cover $\{f_\alpha\}$ in A there is an $f \in A$ such that $f_\alpha f f_\alpha = f_\alpha$ for all α.

The condition of $\mathscr{S}$-regularity provides the means by which we avoid order assumptions in the characterizations of $C(X)$. In general, however, it is not

[4] If A is a subring of $F(X)$, then $\mathfrak{a}(I)$ as defined here coincides with $\mathfrak{a}(I)$ as defined in §1.

suitable for a characterization of $M(X, \mathfrak{B})$. For example, let X be uncountable and let $\mathfrak{B}$ be the algebra of countable sets and their complements. If $\mathscr{S}$ is the set of maximal fixed ideals of the algebra $M(X, \mathfrak{B})$, then $M(X, \mathfrak{B})$ is not $\mathscr{S}$-regular. In fact, there is no algebra $M(X, \mathfrak{B}) \subseteq A \subseteq F(X)$ other than $F(X)$ itself which is $\mathscr{S}$-regular relative to its set $\mathscr{S}$ of maximal fixed ideals.

THEOREM 4.1. *Let A be an algebra and let $\mathscr{S}$ be the family of σ-closed real ideals*[5] *of A. Then A is isomorphic to the algebra $C(X)$ for some P-space X if and only if A is $\mathscr{S}$-regular, $\cap \mathscr{S} = 0$, and each $\mathscr{S}$-complemented ideal of A is $\mathfrak{a}$-principal.*

Proof. Let X be a P-space. Again we may assume that X is also a Q-space; hence every real ideal of $C(X)$ is fixed. As before one easily proves that each such ideal is σ-closed. Thus, clearly, $\cap \mathscr{S} = 0$. If $\{f_\alpha\} \subseteq C(X)$ is a discrete $\mathscr{S}$-cover, then the family $\{X - Z(f_\alpha)\}$ is a disjoint open cover of X; hence $f = \Sigma_\alpha f_\alpha$ is in $C(X)$. Since $C(X)$ is regular, there is an $f' \in C(X)$ with $f'f^2 = f$. Now an obvious pointwise argument shows that $f'f_\alpha^2 = f_\alpha$ for each α; therefore $C(X)$ is $\mathscr{S}$-regular. That $C(X)$ satisfies the final condition follows from Lemma 1.1.

Conversely, let A satisfy the stated conditions. Then, as a subdirect sum of fields, A is commutative. Since the zero ideal of A is $\mathscr{S}$-complemented, there is an $f \in A$ such that $\{0\} = \mathfrak{a}(f)$. If $f \in N$ for some $N \in \mathscr{S}$, then, since N is σ-closed, there is a $g \in N$ with $fg = f$. Let $h \in A$; then $f(h - gh) = 0$. But $\{0\} = \mathfrak{a}(f)$, so we have $h = gh \in N$; that is, $A = N$. This contradiction shows that $\{f\}$ is a discrete $\mathscr{S}$-cover. Then since A is $\mathscr{S}$-regular, $f'f^2 = f$ for some $f' \in A$. Thus $\{0\} = \mathfrak{a}(e)$ for some idempotent e $(= f'f)$ in A; in fact, e is easily seen to be an identity for A. Therefore (cf. **(1)**) we may assume that A is (isomorphic to) a subalgebra of $C(X)$ for some completely regular space X and that (i) the maximal fixed subsets of A are the members of $\mathscr{S}$, and (ii) for each $x \in X$ and each neighbourhood U of x, there is an $f \in A$ such that $f(x) = 0$ and $f(y) \geqq 1$ for all $y \notin U$. It therefore remains to prove that X is a P-space and that $A = C(X)$. So let $U = \cap_n U_n$ be a G_δ-set in X, let $x \in U$, and let

$$N_x = \{f \in A; f(x) = 0\} \in \mathscr{S}.$$

By (ii), there is, for each n, an $f_n \in N_x$ such that $f_n(y) \geqq 1$ for all $y \notin U_n$. Since N_x is σ-closed, there is an $f \in N_x$ such that $ff_n = f_n$ for all n. It is clear that $f(x) = 0$ and that $f(y) = 1$ for all $y \notin U$. Consequently U is a neighbourhood of x. This establishes that X is a P-space.

Now let $Z \subseteq X$ be a continuous zero set; that is, Z is closed and open in X. For each $x \in X - Z$, there is an $f \in N_x$ such that $f(Z) = 1$. Therefore $g = 1 - f \in A$ and $g(x) = 1$ and $g(Z) = 0$. We have from this that $I = \{f \in A; Z \subseteq Z(f)\}$ is Z-convex and $\cap Z(I) = Z$. Then with essentially the same argu-

[5] An ideal N of A is *real* if A/N is isomorphic to the real field.

ment as that used in the proof of Lemma 1.1, we conclude that I is $\mathscr{S}$-complemented. Therefore $I = \mathfrak{a}(f)$ for some $f \in A$; thus, using the fact that X is a P-space, it follows that $Z = Z(f)$ for some $f \in A$. Since $X - Z$ is also a continuous zero set, $X - Z = Z(g)$ for some $g \in A$. This clearly implies that $\{f, g\}$ is a discrete $\mathscr{S}$-cover; hence $f'g^2 = g$ for some $f' \in A$. Thus $(f'g)(Z) = 1$ and $(f'g)(X - Z) = 0$. We have proved then that A contains the characteristic function of each continuous zero set of X.

To complete the proof it will suffice to prove that A contains every strictly positive function in $C(X)$, for if $f \in C(X)$, then $f = [(f \vee 0) + 1] - [-(f \wedge 0) + 1]$. So let $f \in C(X)$ be strictly positive and for each positive real number α, set $Z_\alpha = \{x \in X; f(x) = \alpha\}$. Then each Z_α is a continuous zero set; let $e_\alpha \in A$ be the characteristic function of Z_α. Since $\{Z_\alpha\}$ is a disjoint cover of X, it follows that $\{\alpha e_\alpha\}$ is a discrete $\mathscr{S}$-cover in A. Therefore there is an $f' \in A$ such that $(\alpha e_\alpha)^2 f' = \alpha e_\alpha$ for each α. Then for each $x \in Z_\alpha$,

$$f'(x) = \alpha^{-1} = [f(x)]^{-1}.$$

Since $\{Z_\alpha\}$ covers X, it follows that $\{f'\}$ is a discrete $\mathscr{S}$-cover in A. Thus there is an $f'' \in A$ with $(f')^2 f'' = f'$. Clearly then $f'' = (f')^{-1} = f$ and $f \in A$ as desired.

References

1. F. W. Anderson and R. L. Blair, *Characterizations of the algebra of all real-valued continuous functions on a completely regular space*, Illinois J. Math., *3* (1959), 121–133.
2. G. Birkhoff, *Lattice theory* (rev. ed., New York, 1948).
3. G. Birkhoff and R. S. Pierce, *Lattice-ordered rings*, An. Acad. Brasil. Ci., *28* (1956), 41–69.
4. B. Brainerd, *On a class of latttice-ordered rings*, Proc. Amer. Math. Soc., *8* (1957), 673–683.
5. ——— *F-rings of continuous functions I*, Can. J. Math., *11* (1959), 80–86.
6. K. Fan, *Partially ordered additive groups of continuous functions*, Ann. Math., *51* (1950), 409–427.
7. L. Gillman and M. Henriksen, *Concerning rings of continuous functions*, Trans. Amer. Math. Soc., *77* (1954), 340–362.
8. P. R Halmos, *Measure theory* (New York, 1950).
9. E. Hewitt, *Rings of real-valued continuous functions. I*, Trans. Amer. Math. Soc., *64* (1948), 45–99.
10. S. Kakutani, *Concrete representation of abstract (M)-spaces*, Ann. Math., *42* (1941), 994–1024.
11. H. Nakano, *Modern spectral theory* (Tokyo, 1950).
12. J. von Neumann, *Regular rings*, Proc. Nat. Acad. Sci. U.S.A., *22* (1936), 707–713.
13. M. H. Stone, *A general theory of spectra. II*, Proc. Nat. Acad. Sci. U.S.A., *27* (1941), 83–87.

University of Oregon

THE EQUIVALENCE OF TWO EXTREMUM PROBLEMS

J. KIEFER AND J. WOLFOWITZ

1. Introduction. Let $f_1, \ldots, f_k$ be linearly independent real functions on a space X, such that the range R of $(f_1, \ldots, f_k)$ is a compact set in k-dimensional Euclidean space. (This will happen, for example, if the f_i are continuous and X is a compact topological space.) Let S be any Borel field of subsets of X which includes X and all sets which consist of a finite number of points, and let $C = \{\xi\}$ be any class of probability measures on S which includes all probability measures with finite support (that is, which assign probability one to a set consisting of a finite number of points), and which are such that

$$m_{ij}(\xi) = \int_X f_i(x) f_j(x) \xi(dx) \qquad i, j = 1, \ldots, k$$

is defined. In all that follows we consider only probability measures ξ which are in C. Write $M(\xi)$ for the $k \times k$ matrix $||m_{ij}(\xi)||$. When $M(\xi)$ is non-singular, write $[M(\xi)]^{-1} = ||m^{ij}||$. (We shall not always exhibit dependence on ξ.) Letting $f(x)$ denote the column vector with components $f_i(x)$, and letting primes denote transposes, we define

$$d(x; \xi) = f(x)'[M(\xi)]^{-1} f(x)$$

whenever $M(\xi)$ is non-singular.

We consider two extremum problems. The first is to choose ξ so that

(1) $\qquad \xi$ maximizes $\det M(\xi)$.

The second is to choose ξ so that

(2) $\qquad \xi$ minimizes $\max_x d(x; \xi)$.

We also note that the integral with respect to ξ of $d(x; \xi)$ is k; hence, $\max_x d(x; \xi) \geqslant k$, and thus a sufficient condition for ξ to satisfy (2) is

(3) $$\max_x d(x; \xi) = k.$$

The result of this note is that (1), (2), and (3) are equivalent. This result, which seems to have interest *per se*, also strengthens and extends results of the authors **(1)** on the optimum design of regression experiments. A brief description of the connection with the design of such experiments is given below. The proof of the theorem is elementary and brief.

Received March 30, 1959 Research of J. Kiefer was sponsored by the Office of Naval Research. Research of J. Wolfowitz was supported by the United States Air Force under Contract no. AF 18(600)-685 monitored by the Office of Scientific Research.

2. The theorem. For every ξ consider $M(\xi)$ as a point in Euclidean k^2-space, let T be the totality of such points for all ξ in C, and let $\bar{T}$ be the convex closure of T. It is clear that every extreme point of $\bar{T}$ can be achieved by a ξ which assigns probability one to a single point. Since C contains every ξ with finite support, it follows that $T = \bar{T}$. The class C need not, of course, be convex. However, since our argument will be concerned only with the $M(\xi)$, we may argue below as if C were convex. Thus, if ξ_1 and ξ_2 are in C and

$$\frac{\xi_1 + \xi_2}{2}$$

is not, we may still discuss

$$M\left(\frac{\xi_1 + \xi_2}{2}\right),$$

because there exists a ξ in C with finite support, say ξ_3, such that

$$M(\xi_3) = M\left(\frac{\xi_1 + \xi_2}{2}\right).$$

Moreover, if $H - 1$ is the dimension of the linear space spanned by the functions $f_i f_j$, $i \leqslant j$, any $M(\xi)$ is equal to an $M(\xi')$ where the support of ξ' consists of at most H points. This can often be impoved, as in the case where X is the unit interval and $f_i(x) = x^{i-1}$.

Call a subset D of C *linear* if the following condition holds: For every α, $0 \leqslant \alpha \leqslant 1$, and every pair ξ_1, ξ_2 in D, $\alpha\xi_1 + (1 - \alpha)\xi_2$ is in D whenever it is in C. Thus, if C is convex, D is also convex.

We shall prove the following:

THEOREM. *Conditions* (1), (2), *and* (3) *are equivalent. The set* B *of all* ξ *satisfying these conditions is linear, and* $M(\xi)$ *is the same for all* ξ *in* B.

This result has a function space corollary which may be of interest. Suppose ξ satisfies (3) and that Q is a real $k \times k$ matrix such that $QM(\xi)Q'$ is the identity. Then $g = Qf$ is a vector of orthonormal functions with respect to ξ, and $g(x)'g(x) = d(x; \xi)$. Thus we have

COROLLARY. *If* $f_1, \ldots, f_k$ *are linearly independent, continuous, real functions on a compact space* X, *then there is a probability measure* ξ *on* X *and a linear transformation* $g_i = \sum_j a_{ij} f_j$ *such that* $g_1, \ldots, g_k$ *are orthonormal with respect to* ξ *and*

$$\max_x \sum_{i=1}^{k} g_i^2(x) = k.$$

The set of all such ξ *is the set* B *of the theorem.*

Proof of the theorem. We shall say that ξ is a *local solution* of (1) if $\det M(\xi) > 0$ and if, for every ξ',

$$\frac{\partial}{\partial\alpha}\log\det M([1-\alpha]\xi+\alpha\xi')|_{\alpha=0+}\leqslant 0. \tag{4}$$

Now, if $\det M(\xi)>0$, A is such that $AM(\xi)A'$ is the identity, and $AM(\xi')A'$ is diagonal with diagonal elements b_i, then $\det M([1-\alpha]\xi+\alpha\xi')=\det A^{-2}\,\Pi_i[1-\alpha+\alpha b_i]$, from which we easily compute that $-\log\det M([1-\alpha]\xi+\alpha\xi')$ is convex in $\alpha(0<\alpha<1)$ and is strictly convex unless all $b_i=1$ (that is, unless $M(\xi)=M(\xi')$). Hence, if $\det M(\xi')>\det M(\xi)$, equation (4) cannot hold for that ξ'. We conlcude that local solutions of (1) are actual solutions of (1), and of course the converse is true. Moreover, if $\det M(\xi)=\det M(\xi')=h>0$, we have $\det M(\xi/2+\xi'/2)>h$ unless $M(\xi)=M(\xi')$, so that ξ and ξ' cannot both satisfy (1) unless $M(\xi)=M(\xi')$. It follows from this and the linearity in ξ of $M(\xi)$ that, if ξ and ξ' both satisfy (1), then so does $\alpha\xi+(1-\alpha)\xi'$, whenever it is in C.

It now suffices to prove that $\det M(\xi)>0$ and ξ satisfies (4) for all ξ', if and only if ξ satisfies (2), and only if it satisfies (3). First suppose ξ satisfies (4) and that $\det M(\xi)>0$. Performing the differentiation in (4), and denoting by M_{ij} the cofactor of m_{ij}, we have

$$\begin{aligned}0 &\geqslant [\det M(\xi)]^{-1}\sum_{i,j}\frac{\partial\det M}{\partial m_{ij}}\frac{\partial m_{ij}([1-\alpha]\xi+\alpha\xi')}{\partial\alpha}\Bigg|_{\alpha=0}\\ &=[\det M(\xi)]^{-1}\sum_{i,j}\left(\frac{\partial}{\partial m_{ij}}\sum_{q}m_{iq}M_{iq}\right)[m_{ij}(\xi')-m_{ij}(\xi)]\\ &=[\det M(\xi)]^{-1}\sum_{i,j}M_{ij}(\xi)[m_{ij}(\xi')-m_{ij}(\xi)]=\sum_{i,j}m^{ij}(\xi)m_{ij}(\xi')-k.\end{aligned} \tag{5}$$

Letting ξ' give measure one to the point x, we obtain

$$[f(x)]'M(\xi)^{-1}f(x)\leqslant k \tag{6}$$

for all x. Thus, (3) is satisfied and, as we have remarked, this implies (2).

Finally, if (2) is satisfied, we must have (6) for all x, since we have just seen that there always exist ξ's satisfying (3). Hence, for any ξ' with finite support, we obtain $\sum_{i,j}m^{ij}(\xi)m_{ij}(\xi')\leqslant k$. Hence this inequality is valid for all ξ', and (5) is satisfied. This completes the proof of the theorem.

3. Extensions and applications. We remark that it is easy to see that, if R is bounded but not compact, and if $\{\xi_i\}$ is a sequence of measures on S, then $\lim_i\det M(\xi_i)$ is a maximum if and only if $\lim_i\sup_x d(x;\xi_i)$ is a minimum, and if and only if $\lim_i\sup_x d(x;\xi_i)=k$. Similarly, the first part of the corollary holds with the replacement $\sup_x\sum_i g_i^2(x)<k+\epsilon$, for any $\epsilon>0$.

We now describe briefly the statistical applications of the results. An integer N is given, and the statistician must choose N points $x_1,\ldots,x_N$ (not necessarily distinct) corresponding to which he obtains observations on uncorrelated random variables Y_i $(1\leqslant i\leqslant N)$ with common variance σ^2 (perhaps unknown) and with expectation $\sum_{j=1}^{k}\theta_j f_j(x_i)$, where the θ_j are unknown

real parameters. If $\xi(x)$ denotes the proportion of x_i's which are equal to x, we find that the covariance matrix of best linear estimators of $\theta_1, \ldots, \theta_k$ is $N^{-1}\sigma^2[M(\xi)]^{-1}$. The function ξ is called the experiment or the experimental design. A criterion often adopted for choosing a design is to minimize the determinant of the above covariance matrix (the "generalized variance"). Another possible criterion is to minimize the maximum over x of the variance $N^{-1}\sigma^2 d(x;\xi)$ of the "best linear estimator," given ξ, of the "regression function" $\sum \beta_j f_j(x)$. If we consider not merely the class C_N of probability measures ξ which take on only integral multiples of N^{-1} as values, but rather *all* probability measures ξ in C, then our result is that the two optimality criteria are equivalent. Moreover, for any ξ with support on H points which satisfies (1), (2), and (3), there is clearly a ξ' in C_N which achieves (1), (2), and (3) to within a multiplicative factor $1 + 0(N^{-1})$, and is easy to write down from ξ. Since the *exactly* optimum designs are often difficult to obtain, depend on N, and differ for the two criteria, we see the practical importance of our considerations.

It is very helpful to use the interplay of the two criteria (1) and (2) in obtaining a solution. For example, one can sometimes guess that a solution exists which is a member of a class of ξ which depend on several parameters. One may use (1) as the more convenient initial approach, maximize $\det M(\xi)$ over the parametric class, and then verify whether the maximum just obtained is indeed a maximum over *all* ξ (which may be difficult in terms of (1)) by verifying (3). It is useful to note that, if ξ has a set consisting of k points as its support, then it gives equal measure to each of these points. (This is part of Theorem 5 of **(1)**.) Examples which make use of such methods will appear elsewhere, as will generalizations such as one concerned with the minimization of the determinant of a principal minor of $M(\xi)^{-1}$.

Reference

1. J. Kiefer and J. Wolfowitz, *Optimum designs in regression problems*, Ann. Math. Stat., *30* (1959).

Cornell University

ON RELATIVELY INVARIANT MEASURES

MARK MAHOWALD

1. Introduction. In this note we will discuss the question of the measurability of the multiplier function of a relatively invariant measure on a group. That is, for a group G, σ-ring S, and a measure μ defined on the sets of S, we assume: E in S, x in G implies xE is in S and $\mu(xE) = \sigma(x)\mu(E)$ and study the measurability of the function $\sigma(x)$.

The problem was discussed by Halmos **(1**, p. 265**)**, on locally compact groups and there the situation proved to be as nice as it could be, that is, if the measure is a non-trivial, relatively invariant Baire measure then the multiplier function is continuous. We prove two theorems for groups in which no topology is assumed. In the first theorem we assume a shearing condition and answer the question completely. The second theorem places a condition on the measure and weakens the shearing assumption. Its proof is complicated and occupies the major portion of this paper.

2. Definitions and Notation. We shall use the measure-theoretic notation and definitions of **(1)** with these modifications and additions. All measures which are considered are complete.

2.1. *A left-invariant ring*, R, is a ring of subsets of a group, G such that E in R implies xE is in R for all x in G.

2.2. When we say a function, f, is *S-measurable* we mean that for E in S and M a Borel set of the real line, $E \cap f^{-1}(M) \cap N(f)$ is in S. ($N(f) = \{x : f(x) \neq 0\}$.)

2.3. (G, S, μ) will be a measure space such that G is a group and S is a left-invariant σ-ring of subsets.

2.4. If E and xE are measurable and $\mu(xE) = \sigma(x)\mu(E)$ and if μ is not identically equal to zero and is σ-finite then μ is called *relatively invariant* and will be denoted by $(\sigma)\mu$. Note that the definition of $\sigma(x)$ implies that $0 < \sigma(x) < \infty$, all $x \in G, \sigma(xy) = \sigma(x)\sigma(y) = \sigma(yx), \sigma(e) = 1, \sigma(x)\sigma(x^{-1}) = 1$.

2.5. By $H(S)$ we shall mean the hereditary σ-ring generated by S.

2.6. In $(G, H(S), (\sigma)\mu^*)$ we shall define an *outer measure integral* denoted by $\delta^*(E) = \int^*_E f(x)d\mu^*$, where f is an arbitrary non-negative function on G and

$$\delta^*(E) = \lim_{n\to\infty} \sum_{i=1}^{n2^n} (i-1)2^{-n}\mu^*(E_{ni} \cap E)$$

Received March 11, 1959. This paper is part of a thesis submitted to the University of Minnesota. The author is indebted to Professor Gelbaum for his help and guidance during the preparation of this paper. In addition, the author wishes to thank the referee for helpful suggestions, particularly in the proof of Theorem 2.

where $E_{ni} = \{x: (i-1)2^{-n} \leqslant f(x) < i2^{-n} \text{ for } i = 1, \ldots, n2^n\}$. Note that $0 < \sigma(x) < \infty$ implies that if $\delta^* = \int^* \sigma(x^{-1}) d\mu^*$, then $\delta^*(E) = 0$ if and only if $\mu(E) = 0$.

2.7. (G, S) will be said to satisfy the *shearing condition* if the transformation from $G \times G$ to $G \times G$ defined by $\theta(x, y) \to \theta(x, xy)$ is a measurability preserving transformation, (carries $S \times S$ onto $S \times S$).

2.8. By *weak shearing* we shall mean that if $f(x)$ is S-measurable then $g(x, y) = f(xy)$ is $S \times S$-measurable.

2.9. *By condition A* on a measure space we shall mean that the space is the union of a disjoint class $\mathscr{D}$ of measurable sets of finite measure with the property that every measurable set may be covered by countably many sets of $\mathscr{D}$ and a set of measure zero.

Remark. According to Halmos **(1**, p. 132**)** this implies that the Radon–Nikodym theorem is valid.

2.10. We say that (G, S, μ) is *countably coverable* if for every set E of positive measure and any other measurable set F, there exist x_i, $i = 1, 2, \ldots$, such that $F - \cup x_i E$ has measure zero.

Remark. Lebesgue measure is countably coverable.

2.11. By a *measure group* we shall mean a measurable space (G, S) such that G is a group and S is left-invariant and satisfies the shearing condition.

3. Measurability theorems.

THEOREM 1. *Let (G, S) be a measure group and let $(\sigma)\mu$ be a relatively invariant measure defined on S. Then σ is S-measurable.*

Proof. From the definition of shearing we have, for any subset E of $G \times G$, $(\theta(E))_x = xE_x$. (See **(1)**, p. 258.) Let $E = F \times F$, where F is in S. By Fubini's theorem we have that

$$\int \chi_{\theta(E)} d\mu(y) = \mu((\theta(E))_x) = \mu(xE_x) = \sigma(x)\mu(F)\chi_F$$

is a measurable function of x. Therefore, $\sigma(x)\mu(F)\chi_F$ is measurable but $\mu(F)$ is a constant and F is an arbitrary set in S; hence $\sigma(x)$ is S-measurable.

COROLLARY. *In a measure group the existence of one non-trivial measure $(\sigma)\mu$ implies that any other non-trivial $(\sigma')\mu'$ can be written as*

$$\mu'(E) = K \int_E \sigma'/\sigma \, d\mu.$$

Proof. The theorem implies that both σ and σ' are measurable. Let $\theta(E) = \int_E \sigma(x^{-1}) d\mu$ and $\theta'(E) = \int_E \sigma'(x^{-1}) d\mu'$. Both θ and θ' are invariant measures and (G, S, θ) and (G, S, θ') are measurable groups (see **(1)**, page 257). Therefore Theorem 60:B of **(1)** applies and shows that $K\theta = \theta'$. Let

$$f_n = \sum_{m=1}^{M} a_{nm} \chi_{E_{nm}}$$

be a sequence of simple functions monotonically converging to σ'. Then

$$\mu'(E) = \int_E \sigma'(x) d\theta' = \lim_{n\to\infty} \int_E f_n d\theta'$$

$$= \lim_{n\to\infty} K \sum_{m=1}^{M} a_{nm} \int_{E_{nm}\cap E} \sigma(x^{-1}) d\mu = K \int_E (\sigma'/\sigma)\, d\mu.$$

For Theorem 2 we shall need the following lemmas:

LEMMA 1. *For arbitrary non-negative function f on G, δ^*, the outer measure integral of f in $(G, H, (S), \mu^*)$, is an outer measure on $H(S)$ and the σ-ring of μ^*-measurable sets is contained in the σ-ring of δ^*-measurable sets.*

Proof. The fact that δ^* is an outer measure follows immediately from the definition. Let E be μ^*-measurable. Then for arbitrary $A \in H(S)$ we have

$$\delta^*(A) = \lim_{n\to\infty} \sum_{i=1}^{n2^n} (i-1)2^{-n}\mu^*(A \cap E_{ni})$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n2^n} (i-1)2^{-n}[\mu^*(A \cap E_{ni} \cap E) + \mu^*(A \cap E_{ni} \cap E')]$$

$$= \delta^*(A \cap E) + \delta^*(A \cap E').$$

This completes the proof of the lemma.

LEMMA 2. *If $\delta^*(E) = \int^*_E f d\mu^*$, then f is R-measurable, where R is the collection of δ^*-measurable sets.*

Proof. It is sufficient to show E_{Nj} satisfies the Carathéodory criterion for all $A \in H(S)$. For N and j fixed and $n > N$, we have either $E_{ni} \cap E_{Nj} = \phi$ or E_{ni}. Therefore, for arbitrary $A \in H(S)$ we have

$$\delta^*(A) = \lim_{n\to\infty} \sum_{i=1}^{n2^{n-1}} (i-1)2^{-n}(\mu^*(E_{Nj} \cap A \cap E_{ni}) + \mu^*(E'_{Nj} \cap A \cap E_{ni}))$$

$$= \delta^*(A \cap E_{Nj}) + \delta^*(A \cap E'_{Nj}).$$

LEMMA 3. *$\delta^*(E) = \int^*_E \sigma(x^{-1}) d\mu^*$ is an invariant outer measure on $H(S)$ and the restriction of δ^* to S is an invariant measure on S.*

The proof of this lemma is long and will be given in § 4. We now have this

THEOREM 2. *Let $(G, S, (\sigma)\mu)$ satisfy condition A and be countably coverable and suppose that there exists a set $E \in S$ such that $0 < \delta^*(E) < \infty$, (with δ^* as in Lemma 3). Then there exists a σ-ring R containing S and a measure $(\sigma)\mu$ on R which is an extension of μ on S such that σ is R-measurable. If in addition S satisfies the weak shearing condition then $\sigma(x)$ is S-measurable.*

Proof. By Lemma 3, δ^* restricted to S is a measure. Since $\delta^*(E) = 0$ if $\mu(E) = 0$, $\delta^* \ll \mu$ and condition A then implies the Radon–Nikodym theorem is valid. Let f be the $R-N$ derivative. Let $E \in S$ be such that $0 < \delta^*(E) < \infty$. Let A be any set of $\mathscr{D}$. There exist $\{x_i\}$, $i = 1, 2, \ldots, 1$ such that $\cup x_i E \supset A$. Therefore, on A, δ is σ-finite. Hence f can be chosen to be finite-valued on A, hence on G. On each subset F of A such that $\delta^*(F) < \infty$, we have,

$$\delta^*(F) = \int_F f(y)d\mu = \delta^*(xF) = \int_{xF} f(y)d\mu = \sigma(x^{-1}) \int_F f(x^{-1}y)\, d\mu.$$

Therefore, for each x

$$f(y) = \sigma(x^{-1})f(x^{-1}y), \qquad [\mu] \text{ in } y \text{ for } y \text{ in } A. \tag{1}$$

Since the A are disjoint and a countable union of them cover any measurable set to within a set of measure zero the formula is valid for all x, $[y]$, when y is restricted to any measurable set.

(1) implies $\bar{\mu} = \int (f(x))^{-1}d\delta$ is a relatively invariant measure with δ as the multiplier function on the σ-ring of δ^* measurable sets R. Lemma 2 shows that σ is R-measurable. Therefore, we have only to show that $\bar{\mu}$ is an extension of μ. Using Theorem B, page 134 of **(1)**, we have

$$\int_E (f)^{-1}d\delta = \int_E (f)^{-1}fd\mu = \mu(E),$$

for every E in S. Therefore, $\bar{\mu}$ satisfies the theorem and this completes the proof of the first part of the theorem.

The weak shearing condition implies that $f(y)\sigma(x^{-1}) - f(xy) = g(x, y)$ is $R \times R$-measurable. On every set in $R \times R$, $g(x, y)$ is integrable and its integral will be zero by (1) and the Fubini theorem. Let A be any set in D with $\mu(A) > 0$. Then $\delta^*(A) > 0$ and A contains a set of points of positive μ-measure at which $0 < f(y) < \infty$. Let E be the subset of $A \times A$ for which $f(y)\sigma(x^{-1}) - f(xy) \neq 0$. Then $\bar{\mu} \times \bar{\mu}(E) = 0$. Therefore, for almost all y in E,

$$\frac{f(y)}{\sigma(x)} - f(xy) = 0 \qquad [\bar{\mu}].$$

If $A_y = \{x : f(y)\sigma^{-1}(x) - f(xy) = 0\}$, $\bar{\mu}(A_y) = 0$ for almost all y in A by the Fubini theorem. If $\bar{\mu}(A_y) = 0$, then $\delta^*(A_y) = 0$. Whence $\mu(A_y) = 0$, using 2.6. Thus there exists $y \in E$ with $0 < f(y) < \infty$ and such that $f(y)\sigma^{-1}(x) - f(xy) = 0$ for almost all x in $A[\mu]$. The measurability of $f(xy)$ then implies that $\sigma(x)$ is measurable in A, and the definition of A implies that $\sigma(x)$ is S-measurable.

4. Proof of Lemma 3. We shall prove a sequence of remarks which will lead to the lemma.

REMARK 1. $\mu^*(xE) = \sigma(x)\mu^*(E)$ *for all* E *in* $H(S)$.

Proof. This statement is an immediate consequence of the definition of an outer measure and the relative invariance of μ.

In the following let E be any set in $H(S)$ such that $\mu^*(E_{ni} \cap E) < \infty$ for all n and $i \neq 1$.

From Remark 1 we have

$$\delta^*(E) = \lim_{n\to\infty} \sum_{i=1}^{n2^n} (i-1)/2^n \mu^*(E_{ni} \cap E) = \lim_{n\to\infty} \sum_{i=1}^{n2^n} (i-1)/2^n \sigma(y).$$

$$\mu^*(yE_{ni} \cap yE).$$

Let $A(N, i) = \{j: (i-1)/2^n\sigma(y) < (j-1)/2^N < j/2^N < i/2^n\sigma(y)\}$ for $i = 1, \ldots, n2^n$. Note that j in $A(N, i)$ implies $E_{Nj} \subset yE_{ni}$ and that

$$\bigcup_{j\in A(N,i)} E_{Nj} \subset \bigcup_{j\in A(M,i)} E_{Mj}$$

if $N \leqslant M$.

In Remarks 2 and 3 we shall be concerned with a particular i and fixed n and y; hence we shall suppress the i in the notation $A(N)$.

REMARK 2.

$$yE_{ni} = \lim_{N\to\infty} \bigcup_{j\in A(N)} E_{Nj} \cup I$$

where $I = \{x: \sigma(x) = 2^n\sigma(y)/(i-1)\}$.

Proof. From the definition of A we see that the right side is a subset of the left side. Let z be a member of the left side. Then

$$\sigma(z^{-1}) = (i-1)(2^n\sigma(y))^{-1}$$

or

$$(i-1)(2^n\sigma(y))^{-1} < \sigma(z^{-1}) < i(2^n\sigma(y))^{-1}.$$

The first case implies z is in I. For the second case there exists an M and $j \in A(M)$ such that

$$(i-1)/2^n\sigma(y) \leqslant (j-1)2^{-M} < \sigma(z^{-1}) < j/2^M < i/2^n\sigma(y).$$

Therefore z is in the union over $A(M)$ and $\bigcup_{j\in A(N)} E_{Nj}$ is an increasing sequence of sets; hence the remark follows.

REMARK 3. *Let* $a > 0$ *be arbitrary; then, for any* $E \in H(S)$ *such that* $\mu^*(E_{ni} \cap E) < \infty$, *there exists an* M *such that*

$$\mu^*(yE_{ni} \cap yE) \leqslant \sum_{j\in A(M)} \mu^*(E_{Mj} \cap yE) + \mu^*(I \cap yE) + a2^{-n}n^{-3}.$$

Proof. From Remark 2 we have

$$\mu^*(yE_{ni} \cap yE) \leqslant \mu^*(\lim_{N\to\infty} \cup_{j\in A(N)} E_{Nj} \cap yE) + \mu^*(I \cap yE)$$
$$= \lim_{N\to\infty} \mu^*(\cup_{j\in A(N)} E_{Nj} \cap yE) + \mu^*(I \cap yE)$$
$$\leqslant \lim_{N\to\infty} \sum_{j\in A(N)} \mu^*(E_{Nj} \cap yE) + \mu^*(I \cap yE).$$

Since the left side is finite, there exists an M such that the remark holds.

We can do this for each i obtaining an M_i. If we are given y and fix n such that $\sigma(y)n > 1$ and if we let $N_0 = \max\{M_i, \log_2\sigma(y) + n\}$ then Remark 3 holds uniformly in i for all $N \geqslant N_0$. In addition, since $1/2^n\sigma(y) > 1/2^N$, there exists one distinct j_i for each i such that

$$E_{Nj_i} \supset I.$$

We then can prove

REMARK 4.

$$\sum_{i=1}^{n2^n} (i-1)(2^n\sigma(y))^{-1}\mu^*(yE_{ni} \cap yE)$$
$$\leqslant \sum_{j=1}^{N2^N} (j-1)2^{-N}\mu^*(E_{Nj} \cap yE) + \sum_{i=1}^{n2^n} 2^{-N}\mu^*(E_{Nj_i} \cap yE) + a$$

for all $N \geqslant N_0$.

Proof. We shall call the left side of the inequality K_n. Then, from Remark 3 and the definition of $A(N, i)$, we have

$$K_n \leqslant \sum_{i=1}^{n2^n} (i-1)/2^n\sigma(y)\left[\sum_{j\in A(N,i)} \mu^*(E_{Nj} \cap yE) + \mu^*(I \cap yE)\right]$$
$$+ \sum_{i=1}^{n2^n} a(i-1)/n^2 2^{2n}\sigma(y)$$
$$\leqslant \sum_{i=1}^{n2^n} \sum_{j\in A(N,i)} [(j-1)2^{-N}]\mu^*(E_{Nj} \cap yE)$$
$$+ \sum_{i=1}^{n2^n} (i-1)(2^n\sigma(y))^{-1}\mu^*(yE \cap I) + a.$$

Since there exists one j_i for each i, we have for all $N \geqslant N_0$

$$K_n \leqslant \sum_{j=1}^{N2^N} (j-1)2^{-N}\mu^*(E_{Nj} \cap yE)$$
$$+ a + \sum_{i=1}^{n2^n} [(i-1)(2^n\sigma(y))^{-1} - (j_i-1)2^{-N}]\mu^*(E_{Nj_i} \cap yE).$$

Since $(i-1)(2^n\sigma(y))^{-1} - (j_i-1)2^{-N} \leqslant 2^{-N}$, the remark follows.

REMARK 5. *The lemma is true if* $\mu^*(E) < \infty$.

Proof. If $\mu^*(E) = 0$ we are finished. Therefore we shall assume that $0 < \mu^*(E) < \infty$. Then from Remark 4 and the monotonicity of the outer measure, we have

$$K_n \leqslant a + \delta^*(yE) + (n2^n/2^N)\mu^*(yE).$$

Letting $N \to \infty$ we have $K_n \leqslant a + \delta^*(yE)$. This is true for all n from some point on; therefore, $\delta^*(E) \leqslant a + \delta^*(yE)$. Since a is arbitrary we have $\delta^*(E) \leqslant \delta^*(yE)$. Applying this inequality to the set yE and y^{-1} we conclude that $\delta^*(E) \geqslant \delta^*(yE)$ and the remark follows.

REMARK 6. *If $\mu^*(E) = \infty$, then the lemma is true if there exists a K such that $\mu^*(yE_{ni} \cap yE) < 2^n K$ for all n and $i \neq 1$.*

Proof. Since $1/2^N < 1/2^n\sigma(y)$, we have

$$E_{Nj_i} \subset yE_{ni} \cup yE_{n,i-1}.$$

Then from Remark 4 we have

$$K_n \leqslant \delta^*(yE) + \sum_{i=1}^{n2^n} 2^{-N}[\mu^*(yE_{ni} \cap yE) + \mu^*(yE_{n,i-1} \cap yE)] + a$$

$$\leqslant \delta^*(yE) + 2\,2^n K\, n2^n\, 2^{-N} + a.$$

The remark now follows as in Remark 5.

REMARK 7. *If $\mu^*(E) = \infty$ and there does not exist a K as in remark 6, then the lemma is true.*

Proof. Let K' be given. Then there exists an n and $i_0 \neq 1$ such that

$$\mu^*(yE_{ni_0} \cap yE) = \sigma(y)\mu^*(E_{ni_0} \cap E) > 2^n K'\sigma(y).$$

This implies

$$\delta^*(E) \geqslant \sum_{i=1}^{n2^n} (i - 1)2^{-n}\mu^*(E_{ni} \cap E) > K'.$$

Hence $\delta^*(E) = \infty$. The result follows as in Remark 5.

This completes the proof of the Lemma. The case which was excluded just before Remark 2, that is, E such that $\mu^*(E_{ni} \cap E) = \infty$ for some n and $i \neq 1$, is clearly covered in Remark 7.

REFERENCE

1. P. R. Halmos, *Measure theory* (New York: Van Nostrand Co., Inc.).

Xavier University
Cincinnati, Ohio

SUMS OF FUNCTIONS OF DIGITS

B. M. STEWART

1. Introduction. We generalize in several directions a paper by Porges **(2)** who considered the integer $F(A)$ obtained from the positive integer A by taking the sum of the squares of the digits of A. Porges showed that if $A > 99$, then $F(A) < A$, so that under iteration of $F(A)$ all the positive integers are divided into a finite number of classes, called orbits in the terminology of Isaacs **(1)**, each containing a finite cycle. For his $F(A)$ Porges showed there are only two orbits: one with the 1-cycle: $1 \to 1$; and the other with the interesting 8-cycle: $4 \to 16 \to 37 \to 58 \to 89 \to 145 \to 42 \to 20 \to 4$.

Consider the set Z of non-negative integers and choose as a base of enumeration any desired integer $B \geqq 2$ (not necessarily $B = 10$). Then only the "digits" $0, 1, 2, \ldots, B - 1$ are needed, in suitable multiplicity, to represent any A of Z. Suppose there is given an arbitrary function assigning to each digit a the value $P(a)$ in Z. (In Porges' example the special function used is $P(a) = a^2$.) Each A in Z has a unique representation to the base B, hence if $F(A)$ is defined to be the sum of the values of $P(a)$, summed over all the digits of A, then not only is $F(A)$ well-defined, but also $F(A)$ is an integer of Z, so $F(F(A))$ is meaningful and continued iteration is possible.

More precisely, let a amd a_i be restricted to the set $0, 1, 2, \ldots, B - 1$ and let a_i' be restricted to the subset $1, 2, \ldots, B - 1$. Then any integer A in the range $B^k \leqq A < B^{k+1}$, $k > 0$, has a unique representation

$$A = a_k' B^k + \sum_0^{k-1} a_i B^i.$$

After $P(a)$ has been given, we make the definitions

$$F(a) = P(a), \qquad F(A) = P(a_k') + \sum_0^{k-1} P(a_i),$$

and thus obtain the type of function which suggested the title of this paper.

We propose to study the growth of the function $F(A)$ and to exhibit certain regularities in the behaviour of $F(A)$ despite the arbitrariness of $P(a)$. For example, it proves easy to demonstrate (Theorem 1) the *existence* of an integer C such that $F(C) \geqq C$ and $F(A) < A$ for every $A > C$. Then a more detailed analysis is presented, using an auxiliary constant S, to construct an algorithm (Theorem 2) for the *evaluation* of C. As an aid in finding the value of S, certain other constants J and L are introduced and they provide further interesting sidelights (Theorems 3, 4, 5) on the behaviour of $F(A)$.

Received January 2, 1957; in revised form January 5, 1960.

These general results are applied to the special case $P(a) = a^t$ with considerable effectiveness (Theorems 6, 7, 8).

A preliminary study is made of the orbit- and cycle-numbers resulting from the iteration of $F(A)$ and the finiteness of these numbers is assured. The teasing irregularities of these numbers are shown by selected tables.

Finally, a brief section is presented concerning products of functions of digits.

2. Existence of C. If proving the existence of C is the only concern, we may assume merely that $P(z)$ is a complex function for which $P(a)$ is defined for every a. Define $F(A)$ as above.

THEOREM 1. *To any real $\epsilon > 0$ there corresponds an integer $C = C(\epsilon)$ such that $|F(C)| \geqq \epsilon C$ and such that $|F(A)| < \epsilon A$ for every $A > C$.*

Proof. Let P be the maximum value of $|P(a)|$. Since $B^k/(k+1)$ is increasing and unbounded for $k = 0, 1, 2, \ldots$, there exists $K = K(\epsilon, P)$ such that $B^k/(k+1) > P/\epsilon$ when $k > K$. If $B^k \leqq A < B^{k+1}$, then $|F(A)| \leqq (k+1)P < \epsilon B^k \leqq \epsilon A$ for all $k > K$. Also $|F(0)| \geqq 0$. Hence C exists, $0 \leqq C < B^{K+1}$.

In the sequel our intention to study iteration of $F(A)$ leads us to insist that the values of $P(a)$ be in Z and to avoid painful details we discuss only the case $\epsilon = 1$. As an aside, note that by the usual interpolation formula there exists a polynomial $P_1(x)$ with rational coefficients and degree at most $B - 1$ which will take on for the set $\{a\}$ the prescribed values $\{P(a)\}$. However, it may be convenient to use polynomials of degree higher than $B - 1$, but of simpler structure, as in the case $P(a) = a^t$ when $t \geqq B$.

3. Algorithm for C. Let $H(A) = F(A) - A$ and $H_i(a) = P(a) - aB^i$ for $i \geqq 0$. If $B^k \leqq A < B^{k+1}$, then for $k > 0$,

$$H(A) = H_k(a_k') + \sum_0^{k-1} H_i(a_i).$$

The properties defining C when $\epsilon = 1$ may now be restated:

$$H(C) \geqq 0, \quad H(A) < 0 \text{ for every } A > C.$$

Let m_i be the maximum value of a for which $H_i(a)$ is a maximum, and let m_i' be the maximum value of a' for which $H_i(a')$ is a maximum. Then in the range $B^k \leqq A < B^{k+1}$ when $k > 0$, the maximum value U_k of $H(A)$ is given by $U_k = H(M_k)$, where

$$M_k = m_k'B^k + \sum_0^{k-1} m_iB^i;$$

when $k = 0$, $U_0 = H(m_0')$. Define $U_{-1} = H(0)$.

Define the integer S by the conditions $U_S \geqq 0$ and $U_k < 0$ for every $k > S$. The existence of S follows immediately from $U_{-1} = P(0) \geqq 0$ and from

Theorem 1, since $H(A) < 0$ for $A > C$ implies $U_k < 0$ for every $k > K$. Hence $-1 \leqq S \leqq K$. (In the next section we give much improved estimates of S.) These observations establish the following

LEMMA. *If* $S = -1$, $C = 0$. *If* $S \geqq 0$, $B^S \leqq M_S \leqq C < B^{S+1}$.

To determine the exact value of C when S is known and $S \geqq 0$, consider

$$U_S = H_S(m'_S) + \sum_0^{S-1} H_i(m_i).$$

Determine a maximum c_S' such that

$$H_S(c'_S) + U_S - H_S(m'_S) \geqq 0. \tag{1}$$

This selection is possible with $B - 1 \geqq c_S' \geqq m_S' \geqq 1$, for at least the choice $c_S' = m_S'$ makes (1) hold, since $U_S \geqq 0$.

Next (assuming $S > 0$) determine a maximum c_{S-1} such that

$$H_{S-1}(c_{S-1}) + H_S(c'_S) + U_S - H_S(m'_S) - H_{S-1}(m_{S-1}) \geqq 0.$$

This choice is possible with $B - 1 \geqq c_{S-1} \geqq m_{S-1}$, for at least the choice $c_{S-1} = m_{S-1}$ is valid, because of the previous step (1).

Proceed recursively from $i + 1$ to i, $S > i \geqq 0$, choosing a maximum c_i such that

$$\begin{aligned} H_i(c_i) + H_{i+1}(c_{i+1}) + \ldots + H_S(c'_S) + U_S \\ - H_S(m'_S) - \ldots - H_{i+1}(m_{i+1}) - H_i(m_i) \geqq 0. \end{aligned} \tag{2}$$

This choice is possible with $B - 1 \geqq c_i \geqq m_i$, for at least $c_i = m_i$ is a valid choice, because of the previous step in the algorithm.

THEOREM 2. *For* $S \geqq 0$, *let*

$$Q = c'_S B^S + \sum_0^{S-1} c_i B^i.$$

Then $Q = C$.

Proof. When $i = 0$, the inequality (2) shows that $H(Q) \geqq 0$. If $B^k \leqq A < B^{k+1}$ and $k > S$, then $H(A) \leqq U_k < 0$, by the definitions of U_k and S. If every digit of Q is $B - 1$, it follows that $C = Q = B^{S+1} - 1$.

Otherwise, suppose some digits of Q are less than $B - 1$. Then for each

$$A = a'_S B^S + \sum_0^{S-1} a_i B^i$$

in the range $Q < A < B^{S+1}$, there must be an index i, $S \geqq i \geqq 0$, such that either $B - 1 \geqq a_S' > c_S'$; or $a_S' = c_S'$ and $a_j = c_j$ when $j > i$, but $B - 1 \geqq a_i > c_i$.

In the first case, because of the maximum property of $H_i(m_i)$,

$$H(A) \leqq H_S(a'_S) + \sum_0^{S-1} H_i(m_i) = H_S(a'_S) + U_S - H_S(m'_S) < 0,$$

where the last strict inequality follows from $a_s' > c_s'$ and the maximum property of c_s' expressed in (1).

In the second case, because of the maximum property of $H_r(m_r)$,

$$H(A) \leqq H_s(c_s') + \ldots + H_{t+1}(c_{t+1}) + H_t(a_t) + \sum_0^{t-1} H_r(m_r)$$
$$= H_t(a_t) + H_{t+1}(c_{t+1}) + \ldots + H_s(c_s') + U_s - H_s(m_s') - \ldots - H_t(m_t) < 0,$$

where the last strict inequality follows from $a_t > c_t$ and the maximum property of c_t expressed in (2).

Since we have shown $H(Q) \geqq 0$ and $H(A) < 0$ for every $A > Q$, it follows that $Q = C$.

In the following example $B = 4$. The table shows $P(a)$, $H_t(a)$ and U_t with a double underline for $H_t(m_t)$ and, if there is a distinction, a single underline for $H_t(m_t')$. All entries are written in the usual way with base 10.

TABLE I

EXAMPLE 1.		B^t	1	4	16	64	256	1024
a	$P(a)$	t	0	1	2	3	4	5
0	100	$H_t(0)$	100	100	100	100	100	100
1	50	$H_t(1)$	49	46	34	−14	−206	−974
2	200	$H_t(2)$	198	192	168	72	−312	−1848
3	10	$H_t(3)$	7	−2	−38	−182	−758	−3062
		U_t	198	390	558	630	452	−216

With the aid of the later Corollary 5.1, we may see from this table that $S = 4$. Then starting from $M_4 = B^4 + 2B^2 + 2B + 2$, the algorithm of Theorem 2 is the following. Replacing $m_4 = 1$ by $a = 2$ gives $H = 346$, but by $a = 3$ gives $H = -100$, hence $c_4 = 2$. Next, replacing $m_3 = 0$ by $a = 3$ gives $H = 346 - 100 - 182 = 64$, hence $c_3 = 3$. No further replacements are possible: $c_2 = m_2$, $c_1 = m_1$, $c_0 = m_0$. Thus $C = 2B^4 + 3B^3 + 2B^2 + 2B + 2$.

4. Growth properties of $F(A)$. In this section we obtain further properties of $H(A) = F(A) - A$ and since our chief concern is what happens to $H(A)$ as A increases, we describe these as growth properties of $F(A)$.

Let R be the maximum value of $(P(a') - P(0))/a'$.

If $R < 1$, define $J = 0$. If $1 \leqq R$, define J by $B^{J-1} \leqq R < B^J$.

THEOREM 3. *If $i \geqq J$, $m_i = 0$. If $i < J$, $m_i = m_i'$.*

Proof. Note that $H_t(0) - H_t(a') = P(0) - P(a') + a'B^t > 0$ holds if $B^t > R$, hence for $i \geqq J$. But when $i < J$, suppose $R = (P(m') - P(0))/m'$ and note that $H_t(0) - H_t(m') \leqq 0$.

COROLLARY 3.1. $S \geqq J - 1$.

Proof. If $J = 0$, the statement $S \geqq -1$ is trivial. If $J > 0$ and $i < J$, then it follows from Theorem 3 that $H_i(m_i') \geqq H_i(0) = P(0) \geqq 0$. Hence for $k < J$, $U_k \geqq 0$, therefore $S \geqq J - 1$.

COROLLARY 3.2. *There exists an integer J_1 such that for $i \geqq J_1$, $m_i' = 1$; and if $i < J_1$, then $m_i' > 1$.*

Proof. The proof exactly parallels that of Theorem 3, starting with R_1 as the maximum value of $(P(a') - P(1))/(a' - 1)$ for all $a' > 1$, and defining $J_1 = 0$, if $R_1 < 1$; but otherwise, defining J_1 by $B^{J_1 - 1} \leqq R_1 < B^{J_1}$.

Example 1 provides an illustration of these results wherein $J = 3$, $J_1 = 4$.

COROLLARY 3.3. *The following relations hold*:

$$U_{i+1} - U_i = H_{i+1}(m_{i+1}'), \qquad 0 \leqq i < J; \tag{3}$$

$$U_{i+1} - U_i = H_{i+1}(m_{i+1}') + P(0) - H_i(m_i'), \qquad J \leqq i. \tag{4}$$

Proof. In the sums representing U_{i+1} and U_i, the terms with index $j \leqq i - 1$ are the same, hence

$$U_{i+1} - U_i = H_{i+1}(m_{i+1}') + H_i(m_i) - H_i(m_i').$$

When $0 \leqq i < J$, the second part of Theorem 3 shows $H_i(m_i) = H_i(m_i')$ which establishes (3). When $J \leqq i$, the first part of Theorem 3 shows $H_i(m_i) = H_i(0) = P(0)$ which establishes (4).

COROLLARY 3.4. *Let J_2 be the maximum of J and J_1. If $i \geqq J_2$, then* $U_{i+1} - U_i = P(0) - B^i(B - 1)$.

Proof. From $i \geqq J_2 \geqq J$, relation (4) holds. From $i \geqq J_2 \geqq J_1$, Corollary 3.2 shows $m_{i+1}' = m_i' = 1$, hence

$$H_{i+1}(m_{i+1}') - H_i(m_i') = P(1) - B^{i+1} - (P(1) - B^i),$$

thus (4) reduces to the stated form.

THEOREM 4. *For $i \geqq 0$, $B^i(B - 1) \leqq H(m_i') - H_{i+1}(m_{i+1}') \leqq B^i(B - 1)^2$.*

Proof. From the maximum property of $H_i(m_i')$ it follows that

$$H_i(m_i') - H_{i+1}(m_{i+1}') \geqslant H_i(m_{i+1}') - H_{i+1}(m_{i+1}') = m_{i+1}' B^i(B - 1) \geqq B^i(B - 1).$$

From the maximum property of $H_{i+1}(m_{i+1}')$ it follows that

$$H_i(m_i') - H_{i+1}(m_{i+1}') \leqq H_i(m_i') - H_{i+1}(m_i') = m_i' B^i(B - 1) \leqq B^i(B - 1)^2.$$

COROLLARY 4.1. $m_{i+1}' \leqq m_i'$.

Proof. In the displayed steps of the proof of Theorem 4, note that

$$m'_{i+1}B^i(B-1) \leqq H_i(m'_i) - H_{i+1}(m'_{i+1}) \leqq m'_iB^i(B-1).$$

Define L to be the minimum integer such that $U_{L+1} < U_L$ and such that if $J > 0$, then $L \geqq J-1$; but if $J = 0$, then $L \geqq J$.

We appeal to Corollary 3.4, with i sufficiently large, to show that L must exist. (The existence of L may be shown also by the existence of S and by Corollary 3.1, except for the case $J = 0$ and $S = -1$.)

THEOREM 5. *If* $i \geqq L$, then $U_{i+1} < U_i$.

Proof. The proof is by induction on i with the case L serving as the base for the induction. When $i \geqq J+1$, it follows from (4) and Theorem 4 that

$$\begin{aligned} U_{i+1} - U_i &= H_{i+1}(m'_{i+1}) + P(0) - H_i(m'_i) \leqq P(0) - B^i(B-1) \\ &< P(0) - B^{i-1}(B-1)^2 \leqq P(0) + H_i(m'_i) - H_{i-1}(m'_{i-1}) \\ &= U_i - U_{i-1}. \end{aligned}$$

When $J = 0$ this completes the proof, since $i - 1 \geqq L \geqq J$ implies $i \geqq J+1$.

When $J > 0$ the above argument is valid except for the one possibility $i - 1 = L = J - 1$. But then using $P(0) = H_{J-1}(0)$, the second part of Theorem 3, and (3), we may modify the last displayed line to read

$$H_{J-1}(0) + H_J(m'_J) - H_{J-1}(m'_{J-1}) \leqq H_J(m'_J) = U_J - U_{J-1},$$

which completes the proof.

COROLLARY 5.1. *If* $E \geqq L$ *and if* $U_E \geqq 0$ but $U_{E+1} < 0$, *then* $E = S$.

Proof. Theorem 5 shows $U_k \leqq U_{E+1} < 0$ for every $k > E$. Hence $E = S$.

As an application of this corollary note in Example 1 that $L = 3$, $U_4 > 0$, $U_5 < 0$, consequently $S = 4$.

COROLLARY 5.2. *If* $J > 0$ *and* $i < J - 1$, *then* $U_{i+1} \geqq U_i$.

Proof. If $U_0 < U_{-1}$, then $P(m') - m' < P(0)$ implies $R < 1$ and $J = 0$. So the hypothesis $J > 0$ implies $U_0 \geqq U_{-1}$. Since $i < J - 1$, $i + 1 \leqq J - 1$, and $R = (P(m') - P(0))/m' \geqq B^{J-1} \geqq B^{i+1}$ which implies $P(m') - m'B^{i+1} \geqq P(0)$. Then for $J - 1 > i \geqq 0$, relation (3) holds, so that

$$U_{i+1} - U_i = H_{i+1}(m'_{i+1}) \geqq H_{i+1}(m') = P(m') - m'B^{i+1} \geqq P(0) \geqq 0.$$

Corollary 5.2 indicates that when $J > 0$, the condition $L \geqq J - 1$ is necessary if we are to have $U_{L+1} < U_L$. Thus the search for S, initiated in Corollary 3.1 and made explicit in Corollary 5.1, should begin at this point $L \geqq J - 1$.

However, when $J = 0$, the added condition $L \geqq J$ plays a different role. For $J = 0$ implies $R < 1$, hence $U_0 = P(m_0') - m_0' < P(0) = U_{-1}$, but this does not imply $U_1 < U_0$ as the following example shows.

TABLE II

EXAMPLE 2.		B^i	1	4	16	64	256	1024
a	$P(a)$	i	0	1	2	3	4	5
0	100	$H_i(0)$	100	100	100	100	100	100
1	90	$H_i(1)$	89	86	74	26	−166	−934
2	80	$H_i(2)$	78	72	48	−48	−432	−1968
3	70	$H_i(3)$	67	58	22	−122	−698	−3002
		U_i	89	186	274	326	234	−434

In Example 2, $J = 0$ and $U_0 = 89 < 100 = U_{-1}$. However, $L = 3$ and $S = 4$. Starting from $M_4 = B^4$ we find by the algorithm of Theorem 2 that $C = B^4 + 3B^3 + 1$.

5. The case $P(a) = a^t$. If $P(a) = a^t$ where t is a fixed positive integer, there are two trivial cases. If $t = 1$, it is obvious that $C = B - 1$ for every B. If $B = 2$, it is obvious that $C = 1$ for every t.

THEOREM 6. *If $P(a) = a^t$, $t > 1$, $B > 2$, then $0 < J \leqq S \leqq t$ and S may be determined by: $B^S \leqq J(B-1)^t - (B^J - 2) < B^{S+1}$.*

Proof. Since $P(0) = 0$, $R = (B-1)^{t-1}$, so that J is determined by $B^{J-1} < (B-1)^{t-1} < B^J$. The condition $t > 1$ implies $J > 0$. Moreover, $(B-1)^{t-1} < B^{t-1}$, hence $J \leqq t - 1$.

Since $t > 1$, $H_t(x) = x^t - xB^t$ is concave upward for $x > 0$. Hence m_t' is either 1 or $B - 1$. Note that

$$H_t(1) - H_t(B-1) = 1 + (B-2)B^t - (B-1)^t.$$

When $i \leqq J - 1$,

$$(B-2)B^t < (B-1)B^{J-1} < (B-1)^t,$$

so that $m_i' = B - 1$. When $i \geqq J + 1$,

$$(B-2)B^t \geqq (B-2)B^{J+1} > (B-2)(B-1)^t \geqq (B-1)^t,$$

so $m_i' = 1$. From Theorem 3, it follows that $m_i = m_i' = B - 1$ for $i \leqq J - 1$, and $m_i = 0$ for $i \geqq J$. The only undecided case is m_J' which is either 1 or $B - 1$.

From the preceding results

$$(5) \quad \begin{cases} U_{J-1} = \sum_0^{J-1} H_i(B-1) = \sum_0^{J-1} ((B-1)^t - (B-1)B^t) \\ \qquad\qquad = J(B-1)^t - (B^J - 1); \\ U_J = H_J(m_J') + U_{J-1} \geqq H_J(1) + U_{J-1}; \\ U_i = H_i(1) + U_{J-1}, \quad \text{for} \quad i > J. \end{cases}$$

Using (5) and $(B-1)^{t-1} \geqq B^{J-1}+1$, we may show $U_J \geqq 0$ as follows:

$$\begin{aligned} U_J &\geqq H_J(1) + U_{J-1} = J(B-1)^t - 2(B^J-1) \\ &\geqq J(B-1)(B^{J-1}+1) - 2(B^J-1) = B^{J-1}((J-2)B - J) \\ &\qquad + J(B-1) + 2. \end{aligned}$$

If $J = 1$ or 2 the last expression is 0. If $J \geqq 3$, the last expression is positive, for $B > 2$ implies $(J-2)B \geqq J$. Hence $U_J \geqq 0$, so $S \geqq J$ (a bit more than Corollary 3.1).

If $i > t$, then $B^t \geqq B^{t+1}$; and also from $J \leqq t-1$, we have $i > J+1$. We combine these observations with (5) to see that if $i > t$, then

$$\begin{aligned} U_i &= 1 - B^t + J(B-1)^t - (B^J-1) \\ &< (t-1)(B-1)^t - B^{t+1} < (B-1)^t + (t+1)(B-1)^t - B^{t+1} \\ &< ((B-1)^{t+1} + (t+1)(B-1)^t + \ldots + 1) - B^{t+1} \\ &= (B-1+1)^{t+1} - B^{t+1} = 0. \end{aligned}$$

Since $U_i < 0$ for $i > t$, it follows that $S \leqq t$.

In the proof that $S \geqq J$ we showed that $H_J(1) + U_{J-1} \geqq 0$ which implies $B^J \leqq 1 + U_{J-1}$. From (5) we have $U_t = 1 + U_{J-1} - B^t$ when $i > J$, hence we see that S (with $U_S \geqq 0$ and $U_k < 0$ for all $k > S$) is determined by

$$B^S \leqq 1 + U_{J-1} < B^{S+1}.$$

This result together with (5) completes the proof of Theorem 6.

In general, to find C we must next apply the algorithm of Theorem 2. However, in many cases we can say considerably more, as the following theorem indicates.

THEOREM 7. *If* $P(a) = a^t$, $t > 1$, $B > 2$, *then* $C < (t-1)B^t$. *If* $B > T = (1-(1-t^{-1})^{1/t})^{-1}$ (*which includes all* $B \geqq t^2$) *then*

$$C = (t-1)B^t - 1.$$

Proof. From Theorem 6, $S \leqq t$, hence $C < B^{t+1}$. Suppose that $C < (t-1)B^t$ is false. Then $B^{t+1} > C \geqq (t-1)B^t$ implies $B \geqq t$ and

$$C = c_t' B^t + \sum_0^{t-1} c_i B^i$$

with $B - 1 \geqq c_t' \geqq t-1$. But then it follows that

$$\begin{aligned} C &\geqq c_t'(B-1+1)^t + c_{t-1}B^{t-1} \\ &= c_t'((B-1)^t + t(B-1)^{t-1} + \ldots + 1) + c_{t-1}B^{t-1} \\ &> c_t'(B-1)^t + c_t't(B-1)^{t-1} + c_{t-1}B^{t-1} \end{aligned}$$

$$> (t-1)(B-1)^t + (c_t')^t + (c_{t-1})^t$$

$$\geqq (c_t')^t + \sum_0^{t-1} (c_i)^t = F(C).$$

The inequality $C > F(C)$ is a contradiction of one of the defining properties of C. Therefore $C < (t-1)B^t$ is true, as stated in the first part of Theorem 7.

It is natural to ask for $B \geqq t$ whether $Q = (t-1)B^t - 1$ will serve as C. Since $F(Q) = (t-2)^t + t(B-1)^t$, the inequality $F(Q) \geqq Q$ will hold if $t(B-1)^t \geqq (t-1)B^t$. This is readily brought to the form

$$B > T = (1 - (1 - t^{-1})^{1/t})^{-1}.$$

Since $(1-t^{-2})^t > 1 - t^{-1} > (1-t^{-1})^t$, it follows that $t^2 > T > t$. These observations complete the proof of Theorem 7.

In the remaining cases the method of Theorem 2 is available for finding C. At least one general observation can be made about the result.

THEOREM 8. *For $P(a) = a^t, t > 1, B > 2$, C has the property that $c_i = B - 1$ for $i < J$; and either $c_i = B - 1$ or $c_i \leqq t - 2$ for $J \leqq i \leqq S$.*

Proof. Recall from the proof of Theorem 6 that $m_i = m_i' = B - 1$ for $i \leqq J - 1$. Since $c_i \geqq m_i$ it follows that $c_i = B - 1$ for $i \leqq J - 1$.

The rest of the theorem is trivial if $B \leqq t$, and is known from Theorem 7 if $B > T$. In what follows assume $B > t$.

If $S = t$, it follows from $C < (t-1)B^t$, that $c_i \leqq t - 2$. Since $S \leqq t$, it remains to discuss c_i for the cases $J \leqq i \leqq S$ where $i < t$.

Since $i < t < B$, note that

$$\frac{(B-1)^t - (t-1)^t}{(B-t)} = \sum_0^{t-1} (B-1)^j (t-1)^{t-1-j}$$

$$\geqq \sum_0^{t-1} (B-1)^j \binom{t-1}{j}$$

$$= (B - 1 + 1)^{t-1} = B^{t-1} \geqq B^t.$$

Hence $H_i(B-1) = (B-1)^t - (B-1)B^t \geqq (t-1)^t - (t-1)B^t = H_i(t-1)$. Because of the concave upward property of $H_i(x)$ the inequality $H_i(B-1) \geqq H_i(t-1)$ indicates that the choice of c_i in the range $t - 1 \leqq c_i < B - 1$ would be a contradiction of the requirement in the algorithm of Theorem 2 that c_i be maximal satisfying (1) or (2). Consequently c_i must be limited to the values stated in the theorem.

The following tables illustrate Theorems 6, 7, 8 by showing C for $P(a) = a^t$ for all $B \geqq 3$ when $t = 2, 3, 4, 5$.

TABLE III

$t = 2$	
B	C
$B \geqq 3$	$B^2 - 1$

TABLE IV

$t = 3$	
B	C
3	$2B^2 - 1$
4, 5	$B^3 - 1$
6, 7	$B^3 + B^2 - 1$
$B \geqq 8$	$2B^3 - 1$

TABLE V

$t = 4$	
B	C
3	$B^3 - 1$
4, 5	$B^4 - 1$
6	$B^4 + B^3 - 1$
7 to 11	$2B^4 - 1$
12, 13	$2B^4 + B^3 - 1$
14	$2B^4 + 3B^3 - 1$
$B \geqq 15$	$3B^4 - 1$

TABLE VI

$t = 5$	
B	C
3	$B^4 - 1$
4	$B^5 - 1$
5	$B^5 + B^4 - 1$
6, 7, 8	$2B^5 - 1$
9	$2B^5 + B^4 - 1$
10 to 19	$3B^5 - 1$
20	$3B^5 + B^4 - 1$
21	$3B^5 + 2B^4 - 1$
22	$3B^5 + 3B^4 - 1$
$B \geqq 23$	$4B^5 - 1$

The effectiveness of the algorithm for finding C may be illustrated by an example such as $B = 10, t = 100$. The necessary comparisons are in this case successfully made with a table of logarithms.

	Test	*Decision*
(1)	$10^{J-1} < 9^{99} < 10^J$	$J = 95$
(2)	$10^S \leqq U_{94} + 1 = 95 \cdot 9^{100} - 10^{95} + 2 < 10^{S+1}$ (Remember from (5) that $U_{97} = H_{97}(1) + U_{94}$.)	$S = 97$
(3)	$c^{100} - c \cdot 10^{97} + U_{94} \geqq 0$	$c_{97}' = 2$
(4)	$c^{100} - c \cdot 10^{96} + 2^{100} - 2 \cdot 10^{97} + U_{94} \geqq 0$	$c_{96} = 5$
(5)	$c^{100} - c \cdot 10^{95} + 5^{100} - 5 \cdot 10^{96} + 2^{100} - 2 \cdot 10^{97} + U_{94} \geqq 0$	$c_{95} = 1$

Theorem 8 guarantees $c_i = 9$ for $0 \leqq i < J$, so the algorithm closes, and $C = 2 \cdot B^{97} + 5 \cdot B^{96} + B^{95} + (B^{95} - 1)$.

6. Orbits of F-related integers. Return now to the general function $P(a)$ requiring only that $P(a)$ is a non-negative integer. This modest restriction not only allows the number C to be determined as in Theorem 2, but also allows the function $F(A)$ to be iterated.

Define $F^{(0)}(A) = A$ and $F^{(k+1)}(A) = F(F^{(k)}(A))$. Integers X and Y are said to be F-related if and only if there exist non-negative integers k and m such that $F^{(k)}(X) = F^{(m)}(Y)$. Being F-related is an equivalence relation dividing all non-negative integers into N disjoint sets of F-related integers. Following Isaacs **(1)** call each such set an orbit and denote the orbit containing A by $\{A\}$.

THEOREM 9. *For $F(A)$ the number N is finite.*

Proof. The existence of C implies that each orbit $\{A\}$ contains at least one integer K with $K \leqq C$, for otherwise the sequence $F^{(n)}(A)$ for $n = 0, 1, 2, \ldots$ (all of whose members belong to $\{A\}$) would be an infinite decreasing sequence of non-negative integers. The existence of such a K for each orbit $\{A\}$ shows that $1 \leqq N \leqq C + 1$.

COROLLARY 9.1. *At least one orbit must be infinite.*

An improved estimate of the value of N may be obtained by noting that the value of $F(A)$ does not depend on the order of the digits of A. For if A_1 is obtained from A merely by permuting the digits (but keeping $a_k' > 0$, of course), then $F(A_1) = F(A)$. Consequently many numbers less than C are apt to be F-related.

Let C^* be the number of integers A, $1 \leqq A \leqq C$, which can be written

$$A = a_k' B^k + \sum_0^{k-1} a_i B^i, \qquad B - 1 \geqq a_k' \geqq a_{k-1} \geqq \ldots \geqq a_1 \geqq a_0 \geqq 0.$$

Then an improved estimate for N is given by $1 \leqq N \leqq C^* + 1$.

From $C < B^{S+1}$ and properties of the binomial coefficients it follows that

$$C^* \leqq \binom{B+S+1}{S+1} - (S+2).$$

The work of Isaacs shows for the iteration of a much more general function G, that each orbit of G-related numbers has at most one "cycle" and various incoming "branches." The word "cycle" has the usual meaning—namely, for $F(A)$ it will mean the existence of a period number p (minimal and positive) and an initial point q such that

$$F^{(i+p)}(A) = F^{(i)}(A) \text{ for all } i \geqq q.$$

If $F^{(m)}(X) = Y$, $m \geqq 1$, then X is called an "antecedent" of Y. If $m = 1$, X is an "immediate antecedent" of Y. If $X \neq Y$, X is a "proper antecedent" of Y. If $F(X) = U$ is in the cycle part of $\{A\}$, but X itself is not in the cycle, then X and all its antecedents constitute a "branch" of $\{A\}$.

THEOREM 10. *For $F(A)$ each orbit $\{A\}$ has a unique cycle.*

Proof. If the orbit $\{A\}$ is non-cyclic, then for all n sufficiently large $F^{(n)}(A) > C$; however, for such n, $F^{(n+1)}(A) < F^{(n)}(A)$ and a contradiction is reached, for we cannot have an infinite decreasing sequence of integers $> C$. Thus each orbit $\{A\}$ must have a finite cycle.

To show that this cycle depends on $\{A\}$ and not on the representative A, we reproduce Isaacs' proof. Suppose U and U' are both in $\{A\}$ and that each is a member of some cycle of $\{A\}$. The first hypothesis implies the existence of k and m so that $F^{(k)}(U) = F^{(m)}(U') = U''$. The second hypothesis now shows that U'' is in the cycle containing U and also in the cycle containing U'. In other words, $\{A\}$ has only one cycle.

COROLLARY 10.1. *Let W be the maximum value of $F(A)$ for $A \leqq C$. Then the period p of the cycle of $\{A\}$ is bounded by $1 \leqq p \leqq W + 1$.*

Proof. In the proof of Theorem 9 we showed that $\{A\}$ contains at least one K with $K \leqq C$. Then $F^{(n)}(K) \leqq W$ for all $n \geqq 0$. For either $C < F^{(n)}(K) \leqq W$, whence $F^{(n+1)}(K) < F^{(n)}(K)$ by the definition of C, thus $F^{(n+1)}(K) < W$; or $0 \leqq F^{(n)}(K) \leqq C$, whence $F^{(n+1)}(K) \leqq W$ by the definition of W. Not only is the existence of a cycle of $\{A\}$ newly evident, but also the maximum number of elements in the cycle is the complete set $0 \leqq X \leqq W$, hence $1 \leqq p \leqq W+1$.

COROLLARY 10.2. *Each element U of the cycle part of $\{A\}$ has the property $U \leqq W$ and at least one member U satisfies $U \leqq C$.*

A simple example in which the maximums of both N and p are attained is given by $B = 2$, $P(0) = 1$, $P(1) = 0$, wherein $C = 0$, $W = 1$, and there is just one orbit: $N = 1 = C + 1$, with $p = 2 = W + 1$.

There seem to be few additional general statements to be made about the orbits, cycles, and branches, for by varying $P(a)$ properly, we may construct bizarre situations which contradict proposed generalizations.

REMARK 1. *Not every orbit need be infinite.* For if $P(q) = q$, but $P(a) > q$ when $a \neq q$, then $\{q\}$ contains only q.

REMARK 2. *If $P(a) = 1$ for some $a \neq 0$, then every $Y > 1$ has a proper antecedent.* For $F(A) = Y$ has a solution

$$A = a \sum_{0}^{Y-1} B^i$$

and $A > Y$.

REMARK 3. *If $P(a) = 0$ for some a, and if $A \neq 0$, then $F(A)$ has infinitely many immediate antecedents.* For since $P(a) = 0$,

$$A_m = AB^m + a \sum_{0}^{m-1} B^i$$

has $F(A_m) = F(A)$, for $m = 1, 2 \ldots$. And since $A \neq 0$, the A_m are distinct (even if $a = 0$).

REMARK 4. *If $P(a) > 0$ for all a, then each Y has at most a finite number of immediate antecedents.* For note that if x_a denotes the number of digits of A which are equal to a, then $F(A)$ may be written

$$F(A) = \sum_{0}^{B-1} x_a P(a).$$

Then the assumption $P(a) > 0$ for every a and the restrictions $x_a \geqq 0$ mean that $F(A) = Y$ is a linear Diophantine form problem with at most a finite

number of solutions: $x_0, x_1, \ldots, x_{B-1}$. (Of course, there may be *no* solution.) Corresponding to each such solution set there are only a finite number of integers A resulting from permissible permutations of the sets of digits. (*Permissible* means at least one $x_{a'} > 0$ and $a_k' > 0$ where

$$k = \sum_0^{B-1} x_a.)$$

Example 3. Suppose $B = 10$ and $P(0) = P(2) = P(4) = 18$, $P(6) = 8$, $P(8) = 6$, $P(1) = P(3) = P(5) = 5$, $P(7) = 9$, $P(9) = 7$. It is easy to find $J = 0$, $S = 1$, $M_1 = 20$, $C = 27$, $W = 36$. Then $\{1, 3, 5\}$ is a finite orbit with $p = 1$; and $\{6, 8\}$ and $\{7, 9\}$ are finite orbits each with $p = 2$. All other integers belong to either $\{23\}$, $\{26\}$ or $\{27\}$, all of which are infinite orbits, each with $p = 1$. Hence $N = 6$. These results follow from Corollary 10.2 and Remark 4.

Example 4. Suppose $P(a) = q$ for every a. If $0 \leqq q < B/2$, then $N = 1$ with $p = 1$ and $F(C) = C = q$. If $B^S/S \leqq q < B^{S+1}/(S + 2)$, $S \geqq 1$, then $N = 1$ with $p = 1$ and $F(C) = C = (S + 1)q$. If $B^S/(S + 1) \leqq q < B^S/S$, $S \geqq 1$, then $N = 2$ and both orbits have $p = 1$: one (infinite) contains $F(C) = C = (S + 1)q$, the other (finite) contains $F(U) = U = Sq$.

8. Orbits for $P(a) = a^t$. When the previous discussion is applied to the case $P(a) = a^t$, a few additional comments may be made.

Remark 1 applies with $q = 0$. Hence $\{0\}$ contains only 0.

Remark 2 applies with $a = 1$. Hence if $Y > 1$, Y has a proper antecedent. Because $P(0) = 0$, $F(B^t) = 1$, so $Y = 1$ also has a proper antecedent. Note that the orbit $\{1\}$ has $p = 1$.

Remark 3 applies. Hence each $A \neq 0$ has infinitely many immediate antecedents.

Let N_i indicate the number of orbits of F-related integers with period i. Then $N = \sum N_i$.

For $t = 1$ and any B, $N = N_1 = B$. For $C = B - 1$ implies $N \leqq C + 1 = B$ and $P(a) = a$ shows each $\{a\}$ has $p = 1$. Note that the corresponding $F(A) = \sum a_i$ is the function met in arithmetic in the process called "casting-out $(B - 1)$'s" and has the useful property $F(A) \equiv A \bmod B - 1$.

For $B = 2$ and any t, $N = N_1 = 2$. For $C = 1$ shows $N \leqq 2$ and each of $\{0\}$ and $\{1\}$ has $p = 1$. By the same argument $N_1 \geqq 2$ for every t and every B.

If $t = 2$ and B is odd, then N_1 is even and $N_1 \geqq 4$. From Section 5, when $t = 2$, $C = B^2 - 1$, and hence by Corollary 10.2 each 1-cycle must contain either $U = b$ or $U = aB + b$. If $F(b) = b^2 = b$, then $b = 0$ or 1, the cases noted in the previous paragraph. If $F(aB + b) = a^2 + b^2 = aB + b$, then it follows that

$$F((B - a)B + b) = (B - a)^2 + b^2 = B^2 - 2aB + (aB + b) = (B - a)B + b.$$

Also $B - a \neq a$, because B is odd. Hence 1-cycles of this type occur in pairs, thus N_1 is even. Furthermore, at least one choice of a and b is always available: $a = b = (B + 1)/2$. Hence $N_1 \geqq 4$.

Perhaps the best way to show the teasing irregularity of the orbit and cycle numbers of F-related integers when $P(a) = a^t$ is to append the following brief tables.

TABLE VII

$t = 2$

B	N_1	N_2	N_3	Others	N
3	4	1			5
4	2				2
5	4		1		5
6	2			$N_6 = 1$	3
7	6			$N_4 = 2$	8
8	4	2	1		7
9	4	1	1		6
10	2			$N_8 = 1$	3
11	4		2		6
12	4	2	1	$N_{10} = 1$	8
13	8	3			11
14	2	1		$N_9 = 1$	4
15	4	1	3	$N_5 = N_9 = 1 = N_7$	11
16	2			$N_8 = 1$	3

TABLE VIII

$t = 3$

B	N_1	N_2	N_3	Others	N
3	3			$N_4 = 1$	4
4	10				10
5	4		1		5
6	5			$N_3 = 1$	6
7	8	4	2	$N_4 = N_9 = 1$	16
8	7			$N_5 = 1$	8
9	9	2		$N_5 = N_{11} = 1$	13
10	6	2	2		10

Added in proof:

During 1959-60, as part of an NSF Undergraduate Research Project, Joseph C. Ferrar made use of the Michigan State University MISTIC to check and extend Tables VII and VIII. Thanks to this work several corrections have been made in Table VII. The extended tables for $t = 2$ show B from 17 to 32 and for $t = 3$ show B from 11 to 16.

Space allows explanation of just one of these entries.

When $B = 10$ and $t = 3$, then $C = 1999$. From the discussion following Corollary 9.1, there are

$$\binom{12}{3} = 220$$

numbers from 0 to 999 and

$$\binom{11}{3} = 165$$

numbers from 1111 to 1999 which need to be considered. The results are as follows:

$N_1 = 6$: the 1-cycles being 0; 1; 153; 370; 371; 407;

$N_2 = 2$: the 2-cycles being 136, 244, and 919, 1459;

$N_3 = 2$: the 3-cycles being 55, 250, 133, and 160, 217, 352.

Then by Corollary 10.2 each non-negative integer is a member of one and only one of these $N = 10$ orbits.

9. Products of functions of digits. Use the previous notation for a, a' and A and suppose $P(a)$ is a rational integer $P(a) \geqq 0$. Define

$$G(a) = P(a), \qquad G(A) = P(a_k') \prod_0^{k-1} P(a_i).$$

The question suggested by Theorem 1 (for $\epsilon = 1$) is whether there exists an integer D for which $G(D) \geqq D$ and $G(A) < A$ for every $A > D$.

Let M indicate the maximum value of $P(a)$ and let M' indicate the maximum value of $P(a')$.

CASE 1. *If $M' \geqq B$, then D does not exist.*

Proof. If $P(b') = M'$, then

$$A = b' \sum_0^k B^i$$

has $G(A) = (M')^{k+1} \geqq B^{k+1} > A$ for every k.

CASE 2. *If $M' = 0$, then $D = 0$.*

Proof. If $A > 0$, $P(a_k') = 0$, so $G(A) = 0 < A$. And $G(0) = P(0) \geqq 0$.

CASE 3. *If $0 < M' < B$ and $M \geqq B + 1$, then D does not exist.*

Proof. The hypotheses imply $P(0) = M$. Then $A = b'B^k$ has $G(A) = M'M^k \geqq (B + 1)^k > B^{k+1} > A$, for all k sufficiently large.

CASE 4. *If $M < B$, then D exists.*

Proof. Note $M' \leqq M$. If $B^k \leqq A < B^{k+1}$ and if $k \geqq (B - 1)(B - 2) = k_1$ then $G(A) \leqq M'M^k \leqq (B - 1)^{k+1} < B^k \leqq A$. For from the assumption $k \geqq k_1$ it follows that $B - 1 \leqq 1 + k/(B - 1) < (1 + 1/(B - 1))^k = (B/(B - 1))^k$. Since $G(0) \geqq 0$, D exists and is in the range $0 \leqq D < B^{k_1}$.

CASE 5. *If $M' < B$, if $M = B$, and if $P(a') \geqq a'$ for any a', then D does not exist.*

Proof. The hypotheses imply $P(0) = B$. Hence if $A = a'B^k$, then $G(A) = P(a')B^k \geqq a'B^k = A$, for every k.

CASE 6. *If $M \leqq B$ and $P(a') < a'$ for every a', then $D = 0$.*

Proof. If $B^k \leqq A < B^{k+1}$, then

$$G(A) = P(a_k') \prod_0^{k-1} P(a_i) < a_k' B^k \leqq A$$

for every k. Since $G(0) \geqq 0$, $D = 0$.

Since these six cases exhaust the possible situations, the only "interesting" cases (having $D > 0$) arise when $1 \leqq M' \leqq M < B$ and $P(a') \geqq a'$ for at least one a'. For these cases the actual value of D and the orbits of G-related integers and their cycles may be determined by methods similar to those in §§ 3 and 6.

In particular, the choice $P(a) = a^t$ leads to an "interesting" case only when $t = 1$, and then there are $B - 1$ orbits, each infinite and of period 1.

The author thanks the referee for his stimulating criticisms and suggestions.

References

1. R. Isaacs, *Iterates of fractional order*, Can. J. Math., *2* (1950), 409–416.
2. A. Porges, *A set of eight numbers*, Amer. Math. Monthly, *52* (1945), 379–382.

Michigan State University

A LINEAR DIOPHANTINE PROBLEM

S. M. JOHNSON

1. Introduction. Let $a_1, a_2, \ldots, a_t$ be a set of groupwise relatively prime positive integers. Several authors, **(2; 3; 5; 6)**, have determined bounds for the function $F(a_1, \ldots, a_t)$ defined by the property that the equation

$$n = a_1x_1 + a_2x_2 + \ldots + a_tx_t \tag{1}$$

has a solution in positive integers $x_1, \ldots, x_t$ for $n > F(a_1, \ldots, a_t)$. If $F(a_1, \ldots, a_t)$ is a function of this type, it is easy to see that

$$G(a_1, \ldots, a_t) = F(a_1, \ldots, a_t) - a_1 - a_2 - \ldots - a_t \tag{2}$$

is the corresponding function for the solvability of (1) in non-negative x's.

It is well known that a_1a_2 is the best bound for $F(a_1, a_2)$ and $a_1a_2 - a_1 - a_2$ for $G(a_1, a_2)$. Otherwise only in very special cases have the best bounds been found, even for $t = 3$.

In the present paper a symmetric expression is developed for the best bound for $F(a_1, a_2, a_3)$ which solves that problem and gives insight on the general problem for larger values of t. In addition, some relations are developed which may be of interest in themselves.

2. A General Property. For $t \geqslant 2$, let $B(a_1, a_2, \ldots, a_t)$ be the best bound for $F(a_1, a_2, \ldots, a_t)$, that is, B is the maximum number N where

$$N \neq \sum_{i=1}^{t} x_ia_i \qquad \text{for any } x_i > 0. \tag{3}$$

Then note that B is the maximum N from a restricted set of numbers N satisfying both (3) and

$$N + a_i = \sum_{j=1}^{t} y_{ij}a_j, \qquad y_{ij} > 0 \text{ for each } i. \tag{4}$$

since the definition of B implies B satisfies (4). Thus, in particular,

$$N = (y_{11} - 1)a_1 + y_{12}a_2 + \ldots + y_{1t}a_t, \; y_{1j} > 0.$$

But by (3), $y_{11} - 1 \leqslant 0$ so that $y_{11} = 1$ since $y_{11} > 0$. By symmetry we have

THEOREM 1. *For every N satisfying* (3) *and* (4) *there are representations of N for each $i = 1, 2, \ldots, t$ of the form*

$$N = \sum_{\substack{j=1 \\ j \neq i}}^{t} y_{ij}a_j, \qquad y_{ij} > 0, \tag{5}$$

and B is the maximum such N.

Received October 21, 1957; in revised form March 9, 1959.

3. The Case $t = 3$. A reduction formula. We seek an expression for $B = B(a_1, a_2, a_3)$ having the property that (1) is satisfied for $n > B$ but is not satisfied for $n = B$. Let us first reduce the problem to the case of pairwise relatively prime a's.

Let $d_{ij} = (a_i, a_j)$, $a_i = b_i d_{ij} d_{ik}$, so that $(b_1, b_2) = (b_2, b_3) = (b_3, b_1) = 1$. Then we have

THEOREM 2.

$$B(a_1, a_2, a_3) = d_{12}d_{23}d_{31}B(b_1, b_2, b_3). \tag{6}$$

Proof. First we show that if we write $d = d_{12}$, $\bar{b}_1 = d_{13}b_1$, $\bar{b}_2 = d_{23}b_2$ so that $(d, a_3) = (\bar{b}_1, \bar{b}_2) = 1$, then

$$B(d\bar{b}_1, d\bar{b}_2, a_3) = dB(\bar{b}_1, \bar{b}_2, a_3). \tag{7}$$

Suppose that $dB(\bar{b}_1, \bar{b}_2, a_3) = d\bar{b}_1x + d\bar{b}_2y + a_3z$, $x, y, z > 0$. Then since $(d, a_3) = 1$, we must have $z = wd$, $w > 0$, so that $B(\bar{b}_1, \bar{b}_2, a_3) = \bar{b}_1x + \bar{b}_2y + a_3w$, $x, y, w > 0$, a contradiction to the definition of $B(\bar{b}_1, \bar{b}_2, a_3)$. In addition, for any positive integer $m > 0$, we show that

$$dB(\bar{b}_1, \bar{b}_2, a_3) + m = d\bar{b}_1x + d\bar{b}_2y + a_3z, \qquad x, y, z > 0. \tag{8}$$

We apply a result from **(2)**.

LEMMA 1 (Brauer). *Let a and b be relatively prime positive integers. Then every positive integer m divisible neither by a nor by b is representable either in the form*

$$m = au + bv, \qquad u > 0, v > 0, \tag{9}$$

or

$$m = ab - au - bv, \qquad b > u > 0, a > v > 0. \tag{10}$$

Letting $d = a$ and $a_2 = b$ in Lemma 1, if (9) holds, we have

$$\begin{aligned} dB(\bar{b}_1, \bar{b}_2, a_3) + m &= d(B(\bar{b}_1, \bar{b}_2, a_3) + u) + va_3 \\ &= d\bar{b}_1x + d\bar{b}_2y + a_3(dz + v) \end{aligned} \tag{11}$$

by the definition of $B(\bar{b}_1, \bar{b}_2, a_3)$, giving (8).

If (10) holds, we have $0 < u < a_3$, and $0 < v < d$, so that

$$d(B(\bar{b}_1, \bar{b}_2, a_3) + a_3 - u) - va_3 = d\bar{b}_1x + d\bar{b}_2y + (dz - v)a_3, \tag{12}$$

for x, y, and $(dz - v) > 0$, giving (8).

Finally, if $m = ud$, then (8) follows directly. If $m = va_3$, write $m = da_3 + (v - d)a_3$ giving (8). Thus (7) holds. Applying the method of obtaining (7) twice more gives (6) and Theorem 2.

We have thus reduced the problem to where the a's are pairwise relatively prime. For the moment let $a_1 > a_2 > a_3$. If

$$a_1 = ua_2 + va_3, \qquad u, v > 0, \tag{13}$$

then $B(a_1, a_2, a_3) = a_2a_3 + a_1$ as Brauer showed in **(2)**. Otherwise

$$B(a_1, a_2, a_3) < a_ia_j + a_k. \tag{14}$$

4. An expression for $B(a_1, a_2, a_3)$. We develop a symmetric expression for $B(a_1, a_2, a_3)$ for the case of pairwise relatively prime a's where each $a_i \neq xa_j + ya_k$, $x > 0$, $y > 0$. Later we show that this same form of expression gives the general solution for $t = 3$.

DEFINITION. Let L_i = the minimum positive K_i satisfying

$$K_i a_i = v_{ij} a_j + v_{ik} a_k, \qquad v_{ij} \geqslant 0, v_{ik} \geqslant 0, \qquad i = 1, 2, 3. \tag{15}$$

Such a number exists since $B(a_j, a_k) = a_j a_k < Ka_i$ for large K.

THEOREM 3. *Given*

$$(a_1, a_2) = (a_2, a_3) = (a_3, a_1) = 1 \tag{16}$$

and

$$L_i > 1, \qquad i = 1, 2, 3 \tag{17}$$

and

$$L_i a_i = x_{ij} a_j + x_{ik} a_k, \tag{15'}$$

then the x_{ij} are uniquely defined and

$$x_{ij} > 0. \tag{18}$$

Since $L_i > 1$, it follows from (10) and (16) that

$$a_i = a_j a_k - v_{ij} a_j - v_{ik} a_k \tag{19}$$

where $0 < v_{ij} < a_k$, $0 < v_{ik} < a_j$. Thus $v_{ik} a_k + a_i = (a_k - v_{ij}) a_j \geqslant L_i a_j$ and so by symmetry

$$L_j < a_k, \qquad \text{for each } j \neq k. \tag{20}$$

If $x_{ji} = 0$, then $L_i a_i = x_{jk} a_k$ and by (16) $L_j = ma_k$, a contradiction to (20). This gives (18). Also the x_{ij} are uniquely determined since if $L_i a_i = x_{ij} a_j + x_{ik} a_k = z_{ij} a_j + z_{ik} a_k$, then by (16) we have $x_{ij} = z_{ij} + ma_k$ and $x_{ik} = z_{ik} - ma_j$. If $m > 0$, $x_{ij} > a_k$. But then for some $d > 0$, $L_i a_i = (a_k + d) a_j + x_{ik} a_k$ and by (19) we get $(L_i - 1) a_i = (d + v_{ij}) a_j + (x_{ik} + v_{ik}) a_k$, contradicting the definition of L_i. Similarly, for $m < 0$.

For $t = 3$ and (16) and (17) we show that there are just two numbers N with properties (3) and (4) so that B is the larger of these numbers. From (5) such a number N has representations of the form

$$N = y_{ij} a_j + y_{ik} a_k \qquad i = 1, 2, 3. \tag{5'}$$

Next observe that from (18) we have

$$y_{ij} \leqslant L_j \tag{21}$$

since otherwise for some $d_j > 0$ we would have $N = (L_j + d_j) a_j + y_{ik} a_k = x_{ji} a_i + d_j a_j + (x_{jk} + y_{ik}) a_k$, contradicting (3). From (20) and (21) we have

$$y_{kj} < a_k, \qquad y_{kj} < a_i. \tag{22}$$

Next we show that the representations (5′) for N are unique for each i. For otherwise $y_{ki}a_i + y_{kj}a_j = z_{ki}a_i + z_{kj}a_j$ and from (16) and (22), $y_{kj} - z_{kj} = ma_i$, $m \leqslant 0$, and $y_{ki} - z_{ki} = ma$, $m \geqslant 0$, so that $m = 0$ and $y_{kj} = z_{kj}$, etc.

From (5′) and Theorem 1 we now have unique representations of N of the form

$$N = y_{ki}a_i + y_{kj}a_j = y_{ij}a_j + y_{ik}a_k = y_{jk}a_k + y_{ji}a_i.$$

If $y_{kj} = y_{ij}$, then $y_{ki} = ma_k$, contradicting (22). Thus either $y_{kj} < y_{ij}$ or $y_{kj} > y_{ij}$.

Case 1. If

$$y_{kj} < y_{ij} \tag{23}$$

then $y_{ki}a_i = (y_{ij} - y_{kj})a_j + y_{ik}a_k$ so that $y_{ki} \geqslant L_i$. Thus by (21) we have

$$y_{ki} = L_i. \tag{24}$$

Then by (24) and (5′)

$$N = L_ia_i + y_{kj}a_j = y_{ji}a_i + y_{jk}a_k$$

or $(L_i - y_{ji})a_i + y_{kj}a_j = y_{jk}a_k$, where $L_i > y_{ji}$ by (21). If $L_i = y_{ji}$ then $y_{kj} = ma_k$, contradicting (22), so that $L_i - y_{ji} > 0$ and $y_{jk} \geqslant L_k$ by the definition of L_k. But then $y_{jk} = L_k$ by (21). Thus (23) implies that $y_{ki} = L_i$, $y_{jk} = L_k$, and cyclically, $y_{ij} = L_j$. But then by (15′)

$$N = (x_{ij} + y_{kj})a_j + x_{ik}a_k = L_ja_j + y_{ik}a_k$$

and by the uniqueness of these representations and by cyclic permutation of subscripts, we have

$$y_{ik} = x_{ik} \tag{25}$$

and

$$L_j = x_{ij} + x_{kj}. \tag{26}$$

Thus if $y_{kj} < y_{ij}$, we get a unique number N where

$$N = L_ia_i + x_{kj}a_j \tag{27}$$

with cyclic permutations of subscripts.

Case 2. If

$$y_{kj} > y_{ij}, \tag{28}$$

we get another number where by symmetry

$$N' = L_ia_i + x_{jk}a_k \tag{29}$$

with cyclic permutations of subscripts. $N \neq N'$ since otherwise $x_{jk}a_k = x_{kj}a_j$ which implies $x_{jk} \geqslant a_j$, which by (25) contradicts (22). Note that these two

numbers are the only numbers with properties (3) and (4) for (16), (17), and $t = 3$. Since B is the largest number with property (3), it satisfies (4) so that B is the maximum of N and N' and we have

THEOREM 4. *Given* (16) *and* (17), *then for cyclic permutations of subscripts*

$$B(a_1, a_2, a_3) = L_i a_i + \max (x_{ki} a_j, x_{ji} a_k) \tag{30}$$

and (26) *holds.*

Also it is easy to verify that C, the corresponding best bound for $G(a_1, a_2, a_3)$, satisfies

$$C(a_1, a_2, a_3) + a_1 + a_2 + a_3 = B(a_1, a_2, a_3). \tag{31}$$

5. A computing algorithm for L_i and x_{ij}. Thus we have shown that finding B is equivalent to finding the set of positive integers L_i and x_{ij} exhibited in the form of a matrix of detached coefficients of the three equations (15′) as follows:

a_1	a_2	a_3
$-L_1$	x_{12}	x_{13}
x_{21}	$-L_2$	x_{23}
x_{31}	x_{32}	$-L_3$

In order to develop a simple computing algorithm for these numbers, we need the following result.

LEMMA 2. *Given* $(a_1, a_2) = (a_2, a_3) = (a_3, a_1) = 1$, *then any system of integers* $K_i > 1$ *and* $v_{ij} > 0$ (*not necessarily* L_i *and* x_{ij}) *satisfying* (15) *and* (26) $K_i = v_{ji} + v_{ki}$, *implies that*

$$K_i K_j - v_{ij} v_{ji} = v_{ji} v_{ki} + v_{ki} K_j = \lambda a_k \geqslant a_k \tag{32}$$

for some positive integer λ.

If we write

$$v_{jk}(K_i a_i - v_{ij} a_j) = v_{ik} v_{jk} a_k = v_{ik}(K_j a_j - v_{ji} a_i),$$

then

$$(v_{jk} K_i + v_{ik} v_{ji}) a_i = (v_{ik} K_j + v_{jk} v_{ij}) a_j$$

and (32) follows by (16) and (26).

Furthermore, we have

THEOREM 5. *If* (16) *and* (17) *hold, then the* L_i *and* x_{ij} *in Theorem* 4 *are characterized by the equations* (15′) *and* (26), *and*

$$L_i L_j + x_{ij} x_{ji} = a_k, \tag{33}$$

for cyclic permutations of subscripts. That is, $\lambda = 1$ *in* (32).

Proof. Suppose a system of K_i and v_{ij} satisfy (15), (26), and (33) where at least one $K_i > L_i$, the minimum positive integer satisfying (15).

Case 1. If $K_1 = L_1$, $K_2 = L_2$, then $K_3 = L_3$ by (26) and Theorem 3.

Case 2. Suppose $K_1 = L_1$, but $K_2 > L_2$, $K_3 > L_3$.

Then $x_{12} = v_{12}$ and $x_{13} = v_{13}$ by Theorem 3 and by (15), (26), and (33) $a_1 = K_2K_3 - v_{32}v_{23} = K_2K_3 - (K_2 - x_{12})(K_3 - x_{13}) = x_{12}K_3 + x_{13}K_2 - x_{12}x_{13} > x_{12}L_3 + x_{13}L_2 - x_{12}x_{13} = L_2L_3 - x_{32}x_{23} \geqslant a_1$ by (32), a contradiction to the assumption that $K_2 > L_2$, $K_3 > L_3$.

Case 3. If $K_1 > L_1$, $L_2 > K_2$, $K_3 > L_3$, then first observe that either $v_{ij} > x_{ij}$ or $v_{ik} > x_{ik}$, but not both. For suppose $v_{ij} > x_{ij}$ and $v_{ik} > x_{ik}$. By (33) $v_{ij}v_{jk} + K_jv_{ik} = a_i \leqslant x_{ij}x_{jk} + L_jx_{ik}$ by (32). Thus $v_{jk} < x_{jk}$. Similarly $v_{ij}K_k + v_{ik}v_{kj} = a_i \leqslant x_{ij}L_k + x_{ik}x_{kj}$ so that $v_{kj} < x_{kj}$. But then $a_i \leqslant L_jL_k - x_{jk}x_{kj} < K_jK_k - v_{jk}v_{kj} = a_i$, a contradiction.

In addition either $v_{ji} > x_{ji}$ or $v_{ki} > x_{ki}$ but not both. For suppose $v_{ji} > x_{ji}$ and $v_{ki} > x_{ji}$. By the previous remark $v_{jk} < x_{jk}$, $v_{kj} < x_{kj}$, leading to the same contradiction obtained above. Thus either v_{12}, v_{23}, v_{31}, or v_{21}, v_{32}, v_{13} are larger than the corresponding x's. That is $v_{ij} > x_{ij}$ for cyclic permutations of subscripts.

Suppose v_{21}, v_{32}, v_{13} are larger than x_{21}, x_{32}, x_{13} respectively. Then by (26)

$$(K_2 - L_2)a_2 + (x_{23} - v_{23})a_3 = (v_{21} - x_{21})a_1 \geqslant L_1a_1$$

by the definition of L_1. Thus $v_{21} > L_1$ and by cyclic permutation of subscripts $v_{32} > L_2$, $v_{13} > L_3$.

Finally $a_3 \leqslant L_1L_2 - x_{12}x_{21} < L_1L_2 < v_{21}v_{32} < v_{21}v_{32} + K_3v_{31} = a_3$, a contradiction.

Thus $\lambda = 1$ in (32) implies that $K_i = L_i$, $v_{ij} = x_{ij}$.

Conversely, $\lambda = 1$ in (32), for $K_i = L_i$, $v_{ij} = x_{ij}$ etc. By the following computing algorithm we can always find sets of K_i and v_{ij} with $\lambda = 1$ in (32). Thus they are the desired L_i and x_{ij}. Moreover since the x_{ij} are unique by Theorem 3, λ is unique and must equal 1.

The usefulness of Theorem 5 is apparent since it will be easier to find K's and v's satisfying (15), (26), and (33) rather than find minimal solutions to (15).

The algorithm follows. First we solve for any a_k in terms of a_i and a_j; for instance, for $k = 3$, giving

$$v_{21}a_1 - K_3a_2 + a_3 = 0 \tag{34}$$

with $0 < v_{21} < a_2$, $0 < K_3 < a_1$ by (10), easily done for example as in **(4)**.

Next construct

$$-K_1a_1 + v_{12}a_2 + v_{13}a_3 = 0 \tag{35}$$

where

$$v_{13} = \left[\frac{a_1}{K_3}\right], \qquad K_1 = a_2 - v_{21}v_{13}, a_1 = K_3v_{13} + v_{12}$$

so that $\lambda = 1$ in (32). If $K_1 > v_{21}$, then $K_1 = L_1$, $K_2 = L_2$, and L_3 can be found by (26). Then apply Theorem 4 for $B(a_1, a_2, a_3)$. If $K_1 \leqslant v_{21}$, note that $K_1 \nmid v_{21}$. For if $K_1 | v_{21}$, then since $K_1 = a_2 - v_{21}v_{13}$, $K_1 | a_2$. But then in (34) $K_1 | a_3$. Thus $(a_2, a_3) \geqslant K_1 > 1$, by (17) a contradiction.

Therefore if $K_1 < v_{21}$ we can construct another equation

$$(36) \qquad (v_{21} - pK_1)a_1 - (K_2 - pv_{12})a_2 + (1 + pv_{13})a_3 = 0$$

with

$$p = \left[\frac{v_{21}}{K_1}\right].$$

Since $v_{21} - pK_1 > 0$, $K_2' = K_2 - pv_{12}$ forms a smaller value of K_2 in (34).

Note that the pair of equations giving the smallest values of K_1 and K_2 will still give $\lambda = 1$ in (32). At each stage we repeat the above generating of a smaller K_1 or K_2 until eventually $K_1 = L_1$, $K_2 = L_2$. By Theorem 5 this will come about when we obtain equations of the type (34) and (35) with $K_1 > v_{21}$ and $K_2 > v_{12}$.

To illustrate we find $B(137, 251, 256)$. First calculate that

$$a_1 - 75a_2 + 73a_3 = 0.$$

Then by the algorithm we obtain

$$\begin{aligned} 3a_1 + 31a_2 - 32a_3 &= 0, \\ 7a_1 - 13a_2 + 9a_3 &= 0, \\ 17a_1 + 5a_2 - 14a_3 &= 0. \end{aligned}$$

Thus the matrix of detached coefficients is

a_1	a_2	a_3
−24	8	5
7	−13	9
17	5	−14

and $B = 24a_1 + 9a_3 = 5{,}592$.

It should be pointed out that solving for (34) is not always necessary. Many computational short cuts become apparent after some practice. Note that the suggested algorithm is not merely numerical but gives algebraic relations as well, enabling one to solve all previously solved special cases for $t = 3$ by a unified approach. For example, see the end of the next section.

6. Extensions and restatement of basic theorem. Even if $L_3 = 1$, the statement of Theorems 4 and 5 still holds, dropping the minimality condition on the L_i. In this case, $B = a_1a_2 + a_3$, see **(2)**. But the matrix of coefficients is

a_1	a_2	a_3
$-a_2$	a_1	0
$a_2 - x_{31}$	$-a_1 - x_{32}$	1
x_{31}	x_{32}	-1

with $x_{31} < a_2$. Then $\lambda = 1$, so that Theorem 5 gives the same result $a_1a_2 + a_3$.

Next we show that Theorems 4 and 5 hold even though the a_i's are not reduced to a pairwise relatively prime set b_1, b_2, b_3.

We compare the L's and x_{ij}'s associated with a_1, a_2, a_3 with those L''s and x_{ij}''s associated with b_1, b_2, b_3. From (15'), $L_ia_i = x_{ij}a_j + x_{ik}a_k$, we see that $d_{jk}|L_i$, $d_{ij}|x_{ik}$, $d_{ik}|x_{ij}$. Thus, setting $L_i = d_{jk}L_i'$, $x_{ik} = d_{ij}x_{ik}'$, we have

(35) $\qquad L_jL_k - x_{jk}x_{kj} = a_i$ if and only if $L_j'L_k' - x_{jk}'x_{kj}' = b_i$,

since $d_{ij}d_{ik}(L_j'L_k' - x_{jk}'x_{kj}') = d_{ij}d_{ik}b_i = a_i$.

Finally, all these results can be collected in the following form:

THEOREM 6. *For $(a_1, a_2, a_3) = 1$, define B to be the largest number not of the form $xa_1 + ya_2 + za_3$, $x, y, z > 0$. Then for cyclic permutation of subscripts*

$$B = L_ia_i + \max\,(x_{jk}a_k, x_{kj}a_j),$$

where

$$L_ia_i = x_{ij}a_j + x_{ik}a_k, \qquad L_i > 0,\ x_{ij} \geqslant 0,\ x_{ik} \geqslant 0, \quad L_i = x_{ji} + x_{ki}$$

and

$$L_iL_j - x_{ij}x_{ji} = a_k.$$

The L's and x's can be found either by the computing algorithm discussed in §5, modified to solve first for $d_{ij}a_k$ in terms of a_i and a_j, or by first applying Theorem 2.

In conclusion, observe that the special cases previously obtained for $t = 3$ can be derived directly from the results of this paper.

Example. We can extend the results stated in **(5)** for $B(a, a + 1, a + z)$. Write $a = kz - u$, $0 < u < z$, $k > 1$, $z \geqslant 2$. Then for $u < k + 1$ the coefficient matrix is

$a = a_1 = kz - u$	$a_2 = kz - u + 1$	$a_3 = kz - u + z$
$-(z + k - u)$	$z - u$	$k - 1$
$z - 1$	$-z$	1
$k + 1 - u$	u	$-k$

If $u < 1$, then

$$B = L_3a_3 + x_{12}a_2 = \left(\frac{a + u}{z}\right)(a + z) + (z - u)(a + 1).$$

To correspond to the notation of **(5)**, we solve for $C + 1 = B + 1 - \sum a_i$. Then

$$C + 1 = \left(\frac{a+u}{z}\right) a + (z - 2 - u)a.$$

If $u > 1$, then $B = L_2 a_2 + x_{21} a_1 = k(a + z) + (z - 1)a$, and

$$C + 1 = \left[\frac{a+1}{z}\right](a + z) + (z - 3)a$$

since

$$\left(\frac{a+u}{z}\right) = \left[\frac{a+1}{z}\right] + 1.$$

For $t > 3$, Theorem 1 holds and the author has verified that relations analogous to Theorem 4 hold in many cases. However, this will be the subject of a later paper.

References

1. P. T. Bateman, *Remark on a recent note on linear forms*, Amer. Math. Monthly, *65* (1958), 517–518.
2. A. T. Brauer, *On a problem of partitions—I*, Amer. J. Math., *64* (1942), 299–312.
3. A. T. Brauer and B. M. Seelbinder, *On a problem of partitions—II*, Amer. J. Math., *76* (1954), 343–346.
4. R. J. Levit, *A minimum solution for a diophantine equation*, Amer. Math. Monthly, *63* (1956), 646–651.
5. J. B. Roberts, *Note on linear forms*, Proc. Amer. Math. Soc. (1956), 465–469.
6. ——— *On a diophantine problem*, Can. J. Math., *9* (1957), 219–223.

The Rand Corporation
Santa Monica, California

ARITHMETICAL INVERSION FORMULAS

ECKFORD COHEN

1. Introduction. Let n and r be integers, r positive, and define the *core* $\gamma(r)$ of r to be the product of the distinct prime factors of r ($\gamma(1) = 1$). Let $f(n,r)$ be a complex-valued, arithmetical function of n and r. If for all n, $f(n,r) = f((n,r), r)$ then $f(n, r)$ is called an *even* function (mod r), and if $f(n,r) = f(\gamma(n, r), r)$ for all n, $\gamma(n, r) = \gamma((n, r))$, then $f(n, r)$ is said to be a *primitive* function (mod r). Clearly, both classes of functions are subclasses of the periodic functions (mod r), while the primitive functions form a subclass of the even functions (mod r).

In a series of three papers **(3; 5; 6)** the author developed parallel, though interrelated, trigonometric and arithmetical theories of the even and primitive functions (mod r). It was shown **(3,** Theorem 3**)** that $f(n, r)$ is even (mod r) if and only if it possesses a representation of the form

$$f(n, r) = \sum_{d \mid (n, r)} F\left(d, \frac{r}{d}\right), \tag{1.1}$$

and that $f(n, r)$ is primitive (mod r) if and only if it possesses a representation of the form **(5,** Theorem 8**),**

$$f(n, r) = \sum_{\substack{d \mid \gamma(r) \\ (d, n) = 1}} G\left(d, \frac{r}{d}\right). \tag{1.2}$$

It is the purpose of the present paper to develop a purely arithmetical theory of these two classes of functions, built on the unifying idea of arithmetical inversion.

More precisely, the method of the paper is based on two arithmetical inversion principles, the first (Theorem 2.1) relating to the class of all even functions (mod r), while the second (Theorem 2.3) is limited to the primitive functions (mod r). We remark that the first of these two results becomes equivalent (Corollary 2.2) to the ordinary Möbius inversion formula in case $f(n, r)$ is restricted to the subclass of *completely even* functions (mod r), that is, functions satisfying $f(n, r) = f(n', r')$ for all n, n', and all positive r, r' such that $(n, r) = (n', r')$. An analogous result (Corollary 2.5) is proved for the *completely primitive* functions (mod r), that is, functions satisfying $f(n, r) = f(n', r')$ for all n, n' and all positive r, r' such that $\gamma(r)/\gamma(n, r) = \gamma(r')/\gamma(n', r')$.

Received January 16, 1959.

The characterizations (1.1) and (1.2) of the even and primitive functions (mod r) follow as immediate consequences (Theorems 2.2 and 2.4, respectively) of the above-mentioned inversion relations. Moreover, it also follows that the functions, $F(r_1, r_2)$ and $G(r_1, r_2)$, are *uniquely* determined, under appropriate restrictions on the integral variables r_1 and r_2.

Sections 3 and 4 are devoted to proofs of generalizations of three fundamental identities in the arithmetical theory of even functions. These identities are stated as follows. Let $\mu(r)$ denote the Möbius inversion function and $\phi(r)$ the Euler totient; then

$$\chi(n, r) = \sum_{d \mid (n, r)} d\mu\left(\frac{r}{d}\right) = \frac{\phi(r)\mu(\delta)}{\phi(\delta)} = \Phi(n, r), \tag{1.3}$$

where $\delta = r/(n, r)$;

$$\phi(r) \sum_{\substack{d \mid r \\ (d, n)=1}} \frac{d}{\phi(d)} \mu\left(\frac{r}{d}\right) = \mu(r)\chi(n, r); \tag{1.4}$$

$$\sum_{\substack{d \mid r \\ (d, n)=1}} \frac{\mu^2(d)}{\phi(d)} = \frac{r\phi((n, r))}{\phi(r)(n, r)}. \tag{1.5}$$

Formula (1.3) is Hölder's relation **(7)**, which asserts the equality between the Dedekind-von Sterneck function $\Phi(n, r)$ and Kluyver's function $\chi(n, r)$, or equivalently, the arithmetical form of Ramanujan's sum. The identity (1.4) is due to Brauer and Rademacher **(2; 6, § 5)**, while (1.5) is due in the case $n = 1$ to Landau **(8, p. 182)**; for a proof of the extended form (1.5), we mention **(4, Theorem 9)**. In the sequel, these three relations will be referred to as the Hölder, Brauer-Rademacher, and Landau identities, respectively.

In Theorem 3.1 we give a new proof of a generalization of the Landau identity, proved originally in **(5)**. The proof given in this paper is based on the theory of arithmetical inversion. As a consequence of the generalized Landau identity, we obtain in Theorem 3.2 a wide extension of the Brauer-Rademacher identity.

In Theorem 4.1 we give a new proof, based on arithmetical inversion, of a generalization of the Hölder relation, due to Anderson and Apostol **(1)**. The generalized Landau identity is also used in the proof of Theorem 4.1; moreover, a second proof of this identity is included in § 4, preceding the statement of the extended Hölder formula. The results of the paper are illustrated with a special case in § 5.

It is emphasized that the discussion of this paper is independent of the theory of even functions previously developed. We also mention that the results of the present paper remain valid when the field of values, assumed here to be complex, is replaced by an arbitrary field of characteristic 0.

2. Arithmetical inversion of even functions (mod r). We now prove a general inversion principle for the even functions (mod r).

First we recall the characteristic property of $\mu(r)$,

$$\sum_{d|r} \mu(d) = 1 \quad \text{or} \quad 0 \tag{2.1}$$

according as $r = 1$ or $r > 1$.

THEOREM 2.1. *Let r_1, r_2 denote positive integral variables.*

(A) *If $F(r_1, r_2)$ is an arbitrary function of r_1, r_2 and $f(n, r)$ is an even function* (mod r) *defined by*

$$f(n, r) = \sum_{d|(n,r)} F\left(d, \frac{r}{d}\right) \tag{2.2}$$

then $F(r_1, r_2)$ has the form,

$$F(r_1, r_2) = \sum_{d|r_1} f\left(\frac{r_1}{d}, r\right) \mu(d), \qquad r = r_1 r_2. \tag{2.3}$$

(B) *Conversely, if $f(n, r)$ is an arbitrary even function* (mod r) *and $F(r_1, r_2)$ is defined by* (2.3), *then $f(n, r)$ has the form* (2.2).

Proof. (A) Assume first that $f(n, r)$ is defined by (2.2). Then, placing $r = r_1 r_2$ and using (2.1), it follows that

$$\begin{aligned} \sum_{d|r_1} f\left(\frac{r_1}{d}, r\right) \mu(d) &= \sum_{d|r_1} \mu(d) \sum_{D|((r_1/d), r)} F\left(D, \frac{r}{D}\right) \\ &= \sum_{D|r_1} F\left(D, \frac{r}{D}\right) \sum_{d|(r_1/D)} \mu(d) = F(r_1, r_2). \end{aligned}$$

Thus (A) is proved.

(B) Assuming $F(r_1, r_2)$ to be defined by (2.3), we have, again by (2.1),

$$\begin{aligned} \sum_{d|(n,r)} F\left(d, \frac{r}{d}\right) &= \sum_{d|(n,r)} \sum_{D|d} f\left(\frac{d}{D}, r\right) \mu(D) \\ &= \sum_{\substack{d|(n,r) \\ DE=d}} f(E, r)\, \mu(D) = \sum_{E|(n,r)} f(E, r) \sum_{D|((n,r)/E)} \mu(D) = f((n, r), r), \end{aligned}$$

so that by the definition of an even function (mod r), (B) is proved.

We are thus led immediately to a characterization of the class of even functions (mod r).

THEOREM 2.2. *A function $f(n, r)$ is even* (mod r) *if and only if it has a representation of the form* (2.2). *Moreover, the function $F(r_1, r_2)$ is uniquely determined by* (2.3) *for positive values of r_1 and r_2.*

Replacing $F(r_1, r_2)$ by $F(r_1)$ and $f(n, r)$ by $g((n, r))$, we obtain from Theorem 2.1, with $r_2 = 1$, the following inversion formula for the completely even functions (mod r).

COROLLARY 2.1. *If $F(r)$ is a function of a positive integral variable r, and $f(n, r)$ is a completely even function* (mod r) *defined by*

$$f(n, r) = g((n, r)) = \sum_{d|(n,r)} F(d), \tag{2.4}$$

then $F(r)$ has the form

$$F(r) = \sum_{d|r} f\left(\frac{r}{d}, r\right)\mu(d) = \sum_{d|r} g\left(\frac{r}{d}\right)\mu(d). \tag{2.5}$$

Conversely, if $f(n, r) = g((n, r))$ is completely even (mod r), *and $F(r)$ is defined by* (2.5), *then $f(n, r)$ has the form* (2.4).

Replacing (n, r) in (2.4) by r, Corollary 2.1 becomes the ordinary Möbius inversion formula. In fact,

COROLLARY 2.2. *The inversion relation of Theorem* 2.1 *is equivalent to the Möbius inversion formula, provided the class of functions $f(n, r)$ is restricted to the completely even functions* (mod r).

We also have by Corollary 2.1, the following analogue of Theorem 2.2 (cf. **5**, Theorem 4).

COROLLARY 2.3. *A function $f(n, r)$ is completely even* (mod r) *if and only if it is representable in the form* (2.4). *The function $F(r)$ is uniquely determined by* (2.5) *for $r > 0$.*

The following lemmas are needed in the proof of the inversion theorem for the primitive functions (mod r).

Definition. An integer r is said to be *primitive* if r contains no square factors > 1.

LEMMA 2.1. *If $r = r_1r_2$, $e|\gamma(r)$, and r_1 is primitive, then*

$$\sum_{\substack{d|r_2r_1/\gamma(r) \\ (e,\, r/d)=1}} \chi(r_2, d) = \begin{cases} r\mu(r_1)/\gamma(r) & (e = r_1) \\ 0 & (e \neq r_1). \end{cases}$$

LEMMA 2.2. *If $r = r_1r_2$ then*

$$\sum_{\substack{dc=r \\ (c,\, r_2)=1}} \chi(r_1, d) = r_1\mu(r_2). \tag{2.6}$$

LEMMA 2.3. *If r is primitive, $r_2|r$, and $r_1|r_2$, then*

$$\sum_{\substack{d|r \\ (d,n)=1 \\ \delta|r_1,\, d|r_2}} \mu(d) = \begin{cases} \mu\left(\dfrac{r}{(n,r)}\right) & \text{if } r_1 = (n, r), r_2 = r, \\ 0 & \text{otherwise.} \end{cases}$$

In view of the multiplicative property of $\mu(r)$ and $\chi(n, r)$ as functions of r, it is sufficient to verify the above lemmas in the case that r is the power of a prime. The details are omitted.

THEOREM 2.3. *Let r_1, r_2 represent positive integral variables, r_1 primitive.*

(A) *If $G(r_1, r_2)$ is an arbitrary function of r_1, r_2 and $f(n, r)$ is a primitive function* (mod r) *defined by*

$$f(n,r) = \sum_{\substack{d\mid\gamma(r)\\(d,n)=1}} G\left(d, \frac{r}{d}\right) = \sum_{d\mid((\gamma(r))/(\gamma(n,r)))} G\left(d, \frac{r}{d}\right), \tag{2.7}$$

then $G(r_1, r_2)$ has the form,

$$G(r_1,r_2) = \frac{\gamma(r)\mu(r_1)}{r} \sum_{d\mid((rr_1)/(\gamma(r)))} f\left(\frac{r}{d}, r\right)\chi(r_2,d), \qquad r = r_1r_2. \tag{2.8}$$

(B) *Conversely, if $f(n, r)$ is an arbitrary primitive function* (mod r) *and $G(r_1, r_2)$ is defined by* (2.8), *then $f(n, r)$ has the form* (2.7).

Proof. (A) Assume that $f(n, r)$ is defined by (2.7), and let $T(r_1, r_2)$ denote the right member of (2.8). Then

$$\begin{aligned} T(r_1,r_2) &= \frac{\gamma(r)\mu(r_1)}{r} \sum_{d\mid((rr_1)/(\gamma(r)))} \left(\sum_{\substack{e\mid\gamma(r)\\(e,(r/d))=1}} G\left(e, \frac{r}{e}\right)\right) \chi(r_2,d) \\ &= \frac{\gamma(r)\mu(r_1)}{r} \sum_{e\mid\gamma(r)} G\left(e, \frac{r}{e}\right) \sum_{\substack{d\mid((rr_1)/(\gamma(r)))\\(e,(r/d))=1}} \chi(r_2,d). \end{aligned}$$

Application of Lemma 2.1 yields $T(r_1, r_2) = G(r_1, r_2)$, which proves (A).

(B) Assume $G(r_1, r_2)$ to be given in the form (2.8) and denote the right member of (2.7) by $S(n, r)$.

$$\begin{aligned} S(n,r) &= \frac{\gamma(r)}{r} \sum_{\substack{d\mid\gamma(r)\\(d,n)=1}} \mu(d) \sum_{e\mid((dr)/(\gamma(r)))} f\left(\frac{r}{e}, r\right)\chi\left(\frac{r}{d}, e\right) \\ &= \frac{\gamma(r)}{r} \sum_{e\mid r} f\left(\frac{r}{e}, r\right) \sum_{\substack{d\mid\gamma(r)\\(d,n)=1\\((\gamma(r))/d)\mid(r/e)}} \mu(d) \sum_{D\mid((r/d),e)} D\mu\left(\frac{e}{D}\right) \\ &= \frac{\gamma(r)}{r} \sum_{e\mid r} f\left(\frac{r}{e}, r\right) \sum_{D\mid e} D\mu\left(\frac{e}{D}\right) \sum_{\substack{d\delta=\gamma(r)\\(d,n)=1\\d\mid(r/D),\,\delta\mid(r/e)}} \mu(d). \end{aligned}$$

By Lemma 2.3, the innermost sum of the last expression is 0 unless $\gamma(r/e) = \gamma(n, r)$, $\gamma(r/D) = \gamma(r)$, and under these conditions it has the value $\mu(\gamma(r)/\gamma(n, r))$. Moreover, since $f(n, r)$ is primitive (mod r), we must have then $f(r/e, r) = f(\gamma(r/e), r) = f(\gamma(n, r), r) = f(n, r)$, and it therefore follows, with $m = \gamma(r)/\gamma(n, r)$, that

$$S(n,r) = \frac{\gamma(r)\mu(m)f(n,r)}{r} \sum_{\substack{e\mid r\\ \gamma(r/e)=\gamma(n,r)}} \sum_{\substack{D\mid e\\ \gamma(r/D)=\gamma(r)}} D\mu\left(\frac{e}{D}\right).$$

Note that the conditions $e|r$, $\gamma(r/e) = \gamma(n, r)$ are equivalent to the conditions, $e|(r/\gamma(n, r))$, $(r/e, \gamma(r)) = \gamma(n, r)$. Similarly, $\gamma(r/D) = \gamma(r)$ and $D|(r/\gamma(r))$

are equivalent conditions for a divisor D of r. Therefore by definition of $\chi(n,r)$, one obtains

$$S(n,r) = \frac{\gamma(r)\mu(m)f(n,r)}{r} \sum_{\substack{e\delta=r/\gamma(n,r) \\ (\delta,m)=1}} \chi\left(\frac{r}{\gamma(r)}, e\right).$$

Thus by Lemma 2.2,

$$S(n,r) = \frac{\gamma(r)\mu(m)f(n,r)}{r} \cdot \frac{r\mu(m)}{\gamma(r)} = f(n,r).$$

This completes the proof.

As a consequence of Theorem 2.3, we have the following characterization of the class of primitive functions (mod r).

THEOREM 2.4. *A function $f(n,r)$ is primitive* (mod r) *if and only if it has a representation of the form* (2.7). *Moreover, the function $G(r_1, r_2)$ is uniquely determined, provided r_1 and r_2 are positive and r_1 is primitive.*

Corresponding to Corollaries 2.1, 2.2, and 2.3 in the case of the completely even functions (mod r), we deduce from Theorem 2.3 the following analogous properties of the completely primitive functions (mod r).

COROLLARY 2.4. *If $f(n,r) = k(m)$ is a completely primitive function* (mod r), *$m = \gamma(r)/\gamma(n,r)$, then*

$$f(n,r) = \sum_{\substack{d\mid\gamma(r) \\ (d,n)=1}} G(d) \rightleftarrows G(r_1) = \sum_{d\mid r_1} f\left(\frac{r_1}{d}, r_1\right)\mu\left(\frac{r_1}{d}\right), \tag{2.9}$$

where $G(r_1)$ is defined for primitive integers r_1.

Remark. The equivalence in (2.9) is to be interpreted in the same precise sense as Theorem 2.3.

Proof. Place $r = r_1$, $r_2 = 1$ in (2.7) and note that $\chi(1,d) = \mu(d)$. Formula (2.9) may be reformulated as

$$k(m) = \sum_{d\mid m} G(d) \rightleftarrows G(r_1) = \sum_{d\mid r_1} k(d)\,\mu\left(\frac{r_1}{d}\right), \tag{2.10}$$

where r_1 is primitive and m is defined as in Corollary 2.4. Hence one obtains

COROLLARY 2.5. *If r is primitive, then the inversion relation of Theorem* 2.3 *is equivalent to the Möbius inversion formula, provided $f(n,r)$ is restricted to the completely primitive functions* (mod r).

COROLLARY 2.6 (cf. **5**, Theorem 10). *A function $f(n,r)$ is completely primitive* (mod r) *if and only if it is representable in the form* (2.7) *with $G(r_1, r_2) = G(r_1)$. The function $G(r_1)$ is uniquely determined for positive, primitive r_1.*

3. The generalized Landau and Brauer-Rademacher identities. We first introduce some notation. Let $g(r)$ and $h(r)$ be functions of r and define

$$f(n,r)=\sum_{d\mid(n,r)}h(d)g\left(\frac{r}{d}\right)\mu\left(\frac{r}{d}\right),\quad F(r)=f(0,r). \tag{3.1}$$

Definition. A function $f(r)$ is said to be *completely multiplicative* if $f(1)=1$, $f(r_1r_2)=f(r_1)f(r_2)$ for all r_1, r_2.

We now recall two simple lemmas proved in **(5, § 4)**.

LEMMA 3.1. *If $h(r)$ is completely multiplicative, then*

$$F(r)=h\left(\frac{r}{\gamma(r)}\right)F(\gamma(r)). \tag{3.2}$$

LEMMA 3.2. *If $g(r)$ is multiplicative, $h(r)$ is completely multiplicative, and for all primes p, $h(p)\neq 0$, $h(p)\neq g(p)$, then $F(r)\neq 0$ for all r.*

We now prove a theorem which generalizes the Landau identity, (1.5).

THEOREM 3.1 **(5**, Theorem 9**)**. *If $g(r)$ and $h(r)$ satisfy the conditions of Lemma 3.2, then*

$$\sum_{\substack{d\mid r\\(d,n)=1}}\left(\frac{g(d)}{F(d)}\right)\mu^2(d)=\frac{h(r)F((n,r))}{F(r)h((n,r))}. \tag{3.3}$$

Remark. Since $\mu(r)$, $g(r)$, and $h(r)$ are multiplicative, it follows that $F(r)$ is also multiplicative.

Proof. Denote the right member of (3.3) by $J(n,r)$; in view of the non-vanishing of $F(r)$ and $h(r)$, $J(n,r)$ is properly defined. We verify by Lemma 3.1, and the multiplicative property of $h(r)$ and $F(r)$, that

$$J(n,r)=\frac{h(m)}{F(m)}\left(m=\frac{\gamma(r)}{\gamma(n,r)}\right).$$

Hence $J(n,r)$ is completely primitive (mod r) and we may apply Corollary 2.4. In particular, we have

$$J(n,r)=\sum_{\substack{d\mid\gamma(r)\\(d,n)=1}}G(d), \tag{3.4}$$

where, assuming r_1 *primitive*,

$$G(r_1)=\sum_{d\mid r_1}J\left(\frac{r_1}{d},r_1\right)\mu\left(\frac{r_1}{d}\right)=\sum_{d\mid r_1}\frac{h(d)}{F(d)}\mu\left(\frac{r_1}{d}\right). \tag{3.5}$$

Hence, by the multiplicativity of $\mu(r)$ and $F(r)$, and by Lemma 3.2,

$$\begin{aligned}G(r_1)&=\frac{\mu(r_1)}{F(r_1)}\sum_{d\mid r_1}h(d)\mu(d)F\left(\frac{r_1}{d}\right)\\&=\frac{\mu(r_1)}{F(r_1)}\sum_{d\mid r_1}h(d)\mu(d)\sum_{D\delta=(r_1/d)}h(D)g(\delta)\mu(\delta).\end{aligned}$$

The complete multiplicativity of $h(r)$ gives, with $Dd = E$,

$$G(r_1) = \frac{\mu(r_1)}{F(r_1)} \sum_{E \mid r_1} h(E) g\left(\frac{r_1}{E}\right) \mu\left(\frac{r_1}{E}\right) \sum_{d \mid E} \mu(d).$$

Hence by (2.1),

$$G(r_1) = \frac{\mu^2(r_1) g(r_1)}{F(r_1)}. \tag{3.6}$$

By (3.4) and (3.6) the theorem is proved.

COROLLARY 3.1 ($n = 1$). *Under the conditions of the Theorem,*

$$\sum_{d \mid r} \left(\frac{g(d)}{F(d)}\right) \mu^2(d) = \frac{h(r)}{F(r)} = J(1, r). \tag{3.7}$$

Next we prove a generalization of the Brauer-Rademacher identity (1.4).

THEOREM 3.2. *Under the conditions of Lemma* 3.2,

$$F(r) \sum_{\substack{d \mid r \\ (d, n) = 1}} \frac{h(d)}{F(d)} \mu\left(\frac{r}{d}\right) = \mu(r) f(n, r). \tag{3.8}$$

Proof. Denote the left member of (3.8) by $Q(n, r)$. Let r_1 and r_2 be the uniquely determined positive integers such that $r = r_1 r_2$, $\gamma(r_2) = \gamma(n, r)$, $(r_1, r_2) = 1$. Then on the basis of Corollary 3.1 and the multiplicative property of $\mu(r)$,

$$\begin{aligned} Q(n, r) &= F(r)\mu(r_2) \sum_{d \mid r_1} \frac{h(d)}{F(d)} \mu\left(\frac{r_1}{d}\right) \\ &= F(r)\mu(r_2) \sum_{d \mid r_1} \mu\left(\frac{r_1}{d}\right) \sum_{D \mid d} \frac{g(D)\mu^2(D)}{F(D)}. \end{aligned}$$

With $d = DE$, one obtains then

$$Q(n, r) = F(r)\mu(r_2) \sum_{D \mid r_1} \frac{g(D)\mu^2(D)}{F(D)} \sum_{E \mid (r_1/D)} \mu\left(\frac{r_1/D}{E}\right),$$

so that by (2.1) and the multiplicative property of $\mu(r)$ and $F(r)$,

$$Q(n, r) = F(r_2)\mu(r)\mu(r_1)g(r_1). \tag{3.9}$$

By definition of $F(r)$ and the multiplicativity of $\mu(r)$, $g(r)$, it follows that

$$Q(n, r) = \mu(r)\mu(r_1)g(r_1) \sum_{d \mid r_2} h(d) g\left(\frac{r_2}{d}\right) \mu\left(\frac{r_2}{d}\right) = \mu(r) \sum_{d \mid r_2} h(d) g\left(\frac{r}{d}\right) \mu\left(\frac{r}{d}\right).$$

In view of the presence of the factor $\mu(r)$ and the fact that $\gamma(r_2) = \gamma(n, r)$, one obtains then

$$Q(n, r) = \mu(r) \sum_{d \mid (n, r)} h(d) g\left(\frac{r}{d}\right) \mu\left(\frac{r}{d}\right) = \mu(r) f(n, r).$$

The theorem is proved.

4. The generalized Hölder identity and a second proof of the generalized Landau identity. In the proof of the generalized Landau identity (3.3), we used as starting point the right member $J(n, r)$. This was the natural approach, relative to the application of (3.3) in proving Theorem 3.2, because it was $J(n, r)$, with $m = 1$, that arose in the proof of that theorem. We shall also use (3.3) in the proof of the generalized Hölder theorem below. However, in this proof, it is the left member of (3.3) which arises; therefore, it is proper to give another proof of the generalized Landau identity, proceeding from the left side of (3.3).

Second proof of the generalized Landau identity, Theorem 3.1. Denote the left member of (3.3) by $S(n, r)$. We obtain then by the multiplicative property of $F(r)$ and $g(r)$, with $m = \gamma(r)/\gamma(e)$, $e = (n, r)$,

$$\begin{aligned} S(n, r) &= \sum_{d \mid m} \frac{g(d)}{F(d)} = \frac{1}{F(m)} \sum_{d \mid m} g(d) F\left(\frac{m}{d}\right) \\ &= \frac{1}{F(m)} \sum_{d \mid m} g(d) \sum_{D \mid (m/d)} h(D) g\left(\frac{m/d}{D}\right) \mu\left(\frac{m/d}{D}\right) \\ &= \frac{1}{F(m)} \sum_{D \mid m} h(D) g\left(\frac{m}{D}\right) \sum_{d \mid (m/D)} \mu\left(\frac{m/D}{d}\right). \end{aligned}$$

Hence by (2.1) and multiplicativity,

$$S(n, r) = \frac{h(m)}{F(m)} = \frac{h(\gamma(r))F(\gamma(e))}{F(\gamma(r))h(\gamma(e))}. \tag{4.1}$$

Multiplying both numerator and denominator of the last expression in (4.1) by $h(r/\gamma(r))h(e/\gamma(e))$, one obtains, by the complete multiplicativity of $h(r)$ and by 3.2,

$$S(n, r) = \frac{h(r)F(e)}{F(r)h(e)} \qquad (e = (n, r)),$$

which is (3.3). The proof is complete.

We shall need the following lemma in the proof of the generalized Hölder identity.

LEMMA 4.1. *Under the conditions of Lemma* 3.2, *if a and b are positive integers, then*

$$F(ab) = \frac{F(a)F(b)h((a, b))}{F((a, b))}. \tag{4.2}$$

Proof. In view of the multiplicative property of the functions concerned, it suffices to verify (4.2) in case $a = p^t$, $b = p^s$, p prime, $t \geqq s > 0$. Since $h(r)$ is completely multiplicative, it follows that for $q > 0$, $F(p^q) = h^{q-1}(p)(h(p) - g(p))$. Hence by Lemmas 3.1 and 3.2, one deduces, for the above values of a and b,

$$\frac{F(a)F(b)h((a,b))}{F((a,b))} = \frac{F(p^t)F(p^s)h(p^s)}{F(p^s)} = F(p^t)h^s(p)$$
$$= h^{s+s-1}(p)F(p) = F(p^{s+t}) = F(ab).$$

By multiplicativity, the lemma follows for arbitrary values of a and b.

We now prove the following generalizations of Hölder's identity (1.3).

THEOREM 4.1 (**1**, Theorem 2; **5**, Theorem 2). *If $g(r)$ and $h(r)$ satisfy the conditions of Lemma* 3.2, *then*

$$f(n,r) = \frac{F(r)g(\delta)\mu(\delta)}{F(\delta)} \qquad (\delta = r/(n,r)), \tag{4.3}$$

where $f(n,r)$ is defined by (3.1).

Proof. Denote the right member of (4.3) by $T(n,r)$. Evidently $T(n,r)$ is even (mod r). Hence by Theorem 2.1, $T(n,r)$ has the representation,

$$T(n,r) = \sum_{d\mid(n,r)} H\left(d, \frac{r}{d}\right), \tag{4.4}$$

where, with $r = r_1r_2$,

$$H(r_1,r_2) = \sum_{d\mid r_1} T\left(\frac{r_1}{d}, r\right)\mu(d)$$
$$= F(r)\sum_{d\mid r_1}\frac{g(r_2d)\mu(r_2d)\mu(d)}{F(r_2d)}.$$

But by definition of $\mu(r)$ and multiplicativity, it follows that

$$H(r_1,r_2) = \frac{F(r)g(r_2)\mu(r_2)}{F(r_2)}\sum_{\substack{d\mid r_1\\(d,r_2)=1}}\left(\frac{g(d)}{F(d)}\right)\mu^2(d).$$

Applying Theorem 3.1 one obtains

$$H(r_1,r_2) = \frac{F(r)g(r_2)\mu(r_2)}{F(r_2)}\cdot\frac{h(r_1)F((r_1,r_2))}{F(r_1)h((r_1,r_2))},$$

so that by Lemma 4.1,

$$H(r_1,r_2) = h(r_1)g(r_2)\mu(r_2). \tag{4.5}$$

The theorem follows from (4.4) and (4.5) and the definition of $f(n,r)$.

Combination of (3.8) and (4.3) yields the following result.

COROLLARY 4.1. *Under the conditions of the Theorem,*

$$\sum_{\substack{d\mid r\\(d,n)=1}}\frac{h(d)}{F(d)}\mu\left(\frac{r}{d}\right) = \frac{\mu(r)g(\delta)\mu(\delta)}{F(\delta)} \qquad \left(\delta = \frac{r}{(n,r)}\right). \tag{4.6}$$

5. A special case. In this section we illustrate the results of §§ 3 and 4 with a particular example. Let $J(r) = \phi_2(r)$ denote the Jordan totient of

rank 2. We recall the following identity proved in (**5**, Corollary 24; $t = 2$, $n = 1$):

$$\sum_{d|r} \frac{\mu(d)\phi(d)}{J(d)} = \frac{r\phi(r)}{J(r)}. \tag{5.1}$$

Placing $h(r) = 1$ and $g(r) = \phi(r)/J(r)$ in (3.3), (3.8), (4.3), and (4.6), respectively, one obtains on the basis of (5.1) the following relations.

$$\sum_{\substack{d|r \\ (d,n)=1}} \frac{\mu^2(d)}{d} = \frac{J(r)}{r\phi(r)} \left(\frac{(n,r)\phi((n,r))}{J((n,r))} \right); \tag{5.2}$$

$$\frac{r\phi(r)}{J(r)} \sum_{\substack{de=r \\ (d,n)=1}} \frac{\mu(e)J(d)}{d\phi(d)} = \mu(r) \sum_{\substack{d|(n,r) \\ de=r}} \frac{\mu(e)\phi(e)}{J(e)}; \tag{5.3}$$

$$\sum_{\substack{d|(n,r) \\ de=r}} \frac{\mu(e)\phi(e)}{J(e)} = \frac{\phi(r)(n,r)}{J(r)} \mu\left(\frac{r}{(n,r)} \right); \tag{5.4}$$

$$\sum_{\substack{de=r \\ (d,n)=1}} \frac{\mu(e)J(d)}{d\phi(d)} = \frac{\mu(r)(n,r)}{r} \mu\left(\frac{r}{(n,r)} \right). \tag{5.5}$$

References

1. Douglas R. Anderson and F. M. Apostol, *The evaluation of Ramanujan's sum and its generalizations*, Duke Math. J., *20* (1953), 211–216.
2. A. Brauer and H. Rademacher, *Aufgabe* 31, Jahresbericht der deutschen Mathematiker-Vereinigung, *35* (1926), 94–95 (supplement).
3. Eckford Cohen, *A class of arithmetical functions*, Proc. Nat. Acad. Sci., *41* (1955), 939–944.
4. ——— *Some totient functions*, Duke Math. J., *23* (1956), 515–523.
5. ——— *Representations of even functions* (mod r), *I. Arithmetical identities*, Duke Math. J. *25* (1958), 401–421.
6. ——— *Representations of even functions* (mod r), *II. Cauchy products*, Duke Math. J., *26* (1959), 165–182.
7. O. Hölder, *Zur Theorie der Kreisteilungsgleichung*, Prace Matematyczno-Fizyczne, *43* (1936), 13–23.
8. Edmund Landau, *Ueber die zahlentheoretische Funktion* $\phi(m)$ *und ihre Beziehung zum Goldbachschen Satz*, Göttinger Nachrichten, (1900), 177–186.

University of Tennessee

THE NUMBER OF k-COLOURED GRAPHS ON LABELLED NODES

R. C. READ

Introduction. By a labelled graph we shall mean a set of "nodes," distinguishable from one another and denoted by $A_1, A_2, \ldots$, and a collection of "edges" viz., pairs of nodes. We say that an edge "joins" the pair of nodes which specifies it. We further stipulate that at most one edge joins any two nodes, and that no edge joins a node to itself.

By a "colouring" of a graph in k colours we shall mean a mapping of the nodes of the graph *onto* a set of k colours $C_1, C_2, \ldots, C_k$ such that no two nodes which are joined by an edge are mapped onto the same colour. A graph so coloured in exactly k colours will be called a k-coloured graph. Since it is usually possible to colour a graph in more than one way, there will, in general, be many k-coloured graphs corresponding to a given graph.

The object of this paper is to derive an expression for the number of labelled k-coloured graphs on a given number of nodes. This is a generalization of a result given by Gilbert **(2, § 1)**. Suppose we are given a set of n nodes $A_1, A_2, \ldots, A_n$, a set of positive non-zero integers $n_1, n_2, \ldots, n_k$ such that $n_1 + n_2 + \ldots + n_k = n$ and a set of integers $e_{\alpha\beta} (\alpha, \beta = 1, 2, \ldots, k)$. We shall count the number of k-coloured graphs on these n nodes which are such that n_α nodes are allocated the colour C_α and $e_{\alpha\beta}$ edges join nodes allocated the colour C_α to nodes allocated the colour C_β, $(\alpha, \beta = 1, 2, \ldots, k)$. We let $E = \sum_{\alpha<\beta} e_{\alpha\beta}$ be the total number of edges.

First allocate the colours to the various nodes. This is possible in

$$\frac{n!}{n_1!n_2! \ldots n_k!}$$

different ways. Next consider the number of ways of choosing $e_{\alpha\beta}$ edges joining nodes coloured in C_α and C_β. There are $n_\alpha n_\beta$ possible edges, so the choice can be made in

$$\binom{n_\alpha n_\beta}{e_{\alpha\beta}}$$

different ways. Thus the total number of graphs is

$$\frac{n!}{n_1!n_2! \ldots n_k!} \prod_{\alpha<\beta} \binom{n_\alpha n_\beta}{e_{\alpha\beta}}. \tag{1}$$

Received April 4, 1959.

To find the number of graphs having E edges we must sum expression (1) over all sets $\{e_{\alpha\beta}\}$ such that $\sum e_{\alpha\beta} = E$. Since

$$\binom{n_\alpha n_\beta}{e_{\alpha\beta}}$$

is the coefficient of

$$t^{e_{\alpha\beta}} \quad \text{in} \quad (1+t)^{n_\alpha n_\beta},$$

this sum is the coefficient of t^E in

$$\frac{n!}{n_1!n_2!\ldots n_k!}\prod_{\alpha<\beta}(1+t)^{n_\alpha n_\beta} = \frac{n!}{n_1!n_2!\ldots n_k!}(1+t)^{\sum n_\alpha n_\beta}$$

and is therefore

$$\frac{n!}{n_1!n_2!\ldots n_k!}\binom{\frac{1}{2}n^2 - \frac{1}{2}\sum n_\alpha^2}{E} \tag{2}$$

since

$$\sum n_\alpha n_\beta = \tfrac{1}{2}(\sum n_\alpha)^2 - \tfrac{1}{2}\sum n_\alpha^2.$$

In the special case when there are n colours and each node receives a different colour we have $n_1 = n_2 = \ldots = n_n = 1$ and (2) reduces to

$$n!\binom{\frac{1}{2}n(n-1)}{E}.$$

Since, under these conditions, every graph on n nodes can be coloured in $n!$ different ways, the number of graphs on n labelled nodes having E edges is seen to be

$$\binom{\frac{1}{2}n(n-1)}{E}.$$

This result is easily obtained directly (**2**, p. 405).

To find the total number of k-coloured graphs on n nodes and E edges we need to sum (2) over all sets $\{n_\alpha\}$ such that $\sum n_\alpha = n$. Thus we obtain

$$\sum_{(n)} \frac{n!}{n_1!n_2!\ldots n_k!}\binom{\frac{1}{2}n^2 - \frac{1}{2}\sum n_\alpha^2}{E}$$

but it does not appear that this formula is very amenable to manipulation.

Let us now remove the restriction on the number of edges in the graph, and consider the number of k-coloured graphs (whatever the number of edges) which are associated in the above way with the set $\{n_\alpha\}$. This number is obtained by summing (2) for all possible values of E, and is thus

$$\frac{n!}{n_1!n_2!\ldots n_k!}\,2^{\frac{1}{2}n^2 - \frac{1}{2}\sum n_\alpha^2}. \tag{3}$$

The total number of k-coloured graphs on n labelled nodes can now be found. We denote it by $F_n(k)$, and we see that

$$F_n(k) = \sum_{(n)} \frac{n!}{n_1!n_2!\dots n_n!} 2^{\frac{1}{2}n^2 - \frac{1}{2}\sum n_a^2}$$

$$= n!2^{\frac{1}{2}n^2} \text{ times the coefficient of } x^n \text{ in } \left[\sum_{s=1}^{\infty} \frac{2^{-\frac{1}{2}s^2}}{s!} x^s\right]^k.$$

Hence

$$\sum_{n=1}^{\infty} 2^{-\frac{1}{2}n^2} F_n(k) \frac{x^n}{n!} = \left[\sum_{s=1}^{\infty} \frac{2^{-\frac{1}{2}s^2}}{s!} x^s\right]^k \tag{4}$$

from which $F_n(k)$ may be calculated.

2. For convenient calculation of $F_n(k)$ we may write (4) as

$$\sum_{n=1}^{\infty} 2^{-\frac{1}{2}n^2} F_n(k) \frac{x^n}{n!} = \left(\sum_{r=1}^{\infty} 2^{-\frac{1}{2}r^2} F_r(k-1) \frac{x^r}{r!}\right)\left(\sum_{s=1}^{\infty} 2^{-\frac{1}{2}s^2} \frac{x^s}{s!}\right)$$

whence, equating coefficients of x^n, we obtain

$$F_n(k) = \sum_{r=1}^{n-1} \binom{n}{r} 2^{r(n-r)} F_r(k-1) \tag{5}$$

which gives the numbers of k-coloured graphs in terms of the numbers of $(k-1)$-coloured graphs. Some values of $F_n(k)$ are given in Table I.

TABLE I

n \ k	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0
2	1	4	0	0	0	0	0
3	1	24	48	0	0	0	0
4	1	160	1152	1536	0	0	0
5	1	1440	30720	122880	122880	0	0
6	1	18304	1152000	10813440	29491200	23592960	0
7	1	330624	65630208	1348730880	7707033600	15854469120	10569646080

3. If we wish to count the total number of graphs coloured in k or fewer colours, we proceed as before but remove the restriction that the n_a's are non-zero, and allow them to be any non-negative integers. Denoting the required number of graphs by $M_n(k)$ we obtain

$$\sum_{n=0}^{\infty} 2^{-\frac{1}{2}n^2} M_n(k) \frac{x^n}{n!} = \left[\sum_{s=0}^{\infty} 2^{-\frac{1}{2}s^2} \frac{x^s}{s!}\right]^k. \tag{6}$$

By the method of § 2 we obtain from (6) the relation

$$M_n(k) = \sum_{r=0}^{n} \binom{n}{r} 2^{r(n-r)} M_r(k-1) \tag{7}$$

with $M_0(k) = 1$.

$M_n(k)$, unlike $F_n(k)$, is a polynomial in k of degree n. This follows either from (7) by mathematical induction, or from the fact that $M_n(k)$ is the sum of the chromatic polynomials* of all graphs on n nodes, each polynomial being counted as many times as there are ways of labelling the corresponding graph. Some values of $M_n(k)$ are given in Table II.

TABLE II

n \ k	1	2	3	4	5	6	7	8	9
1	1	2	3	4	5	6	7	8	9
2	1	6	15	28	45	66	91	120	153
3	1	26	123	340	725	1326	2191	3368	4905
4	1	162	1635	7108	20805	48486	97447	176520	296073
5	1	1442	35043	254404	1058885	3216486	7986727		
6	1	18306	1206915	15531268					
7	1	330626	66622083						

The first four polynomials are

$$M_1(k) = k,$$
$$M_2(k) = 2k^2 - k,$$
$$M_3(k) = 8k^3 - 12k^2 + 5k,$$

and

$$M_4(k) = 64k^4 - 192k^3 + 208k^2 - 79k.$$

4. If $f_n(k)$ denotes the number of *connected* k-coloured graphs, it can be shown by the methods used in **(2)** that

$$\sum_{n=0}^{\infty} F_n(k) \frac{x^n}{n!} = \exp\left\{ \sum_{n=1}^{\infty} f_n(k) \frac{x^n}{n!} \right\} \tag{8}$$

with $F_0(k) = 1$.

Differentiating both sides of (8) and equating coefficients of x^{n-1} we obtain

$$F_n(k) = \sum_{r=1}^{n} \binom{n-1}{r-1} F_{n-r}(k) f_r(k)$$

or

*For the definition of the chromatic polynomial of a graph see **(1)**.

$$f_n(k) = F_n(k) - \sum_{r=1}^{n-1} \binom{n-1}{r-1} F_{n-r}(k) f_r(k) \tag{9}$$

giving $f_n(k)$ in terms of $f_{n-1}(k)$, $f_{n-2}(k)$, when $F_n(k)$, $F_{n-1}(k)$, . . . , are known.

REFERENCES

1. H. Whitney, *A logical expansion in mathematics*, Bull. Amer. Math. Soc., *34* (1932), 339–362.
2. E. N. Gilbert, *Enumeration of labelled graphs*, Can. J. Math., *8* (1956), 405–411.

University College of the West Indies

DISCRETE GROUPS OF MOTIONS

LEON GREENBERG

1. Introduction. This paper deals with the discrete groups of rigid motions of the hyperbolic plane. It is known **(12)** that the finitely generated, orientation-preserving groups have the following presentations:

Generators: $a_1, b_1, \ldots, a_p, b_p, S_1, \ldots, S_d, c_1, \ldots, c_r$.

Defining relations: $k_1 \ldots k_p S_1 \ldots S_d c_1 \ldots c_r = 1$,

$$S_1^{n_1} = S_2^{n_2} = \ldots = S_d^{n_d} = 1,$$

where $k_m = a_m b_m a_m^{-1} b_m^{-1}$. We shall denote this group by $F(p; n_1, \ldots, n_d; r)$.

In particular, the finitely generated free groups are contained among these. Indeed, one purpose of this paper is to indicate some geometrical methods for investigating free groups.

The above groups also include the orientation-preserving discrete groups of motions of the sphere and Euclidean plane **(3)**. But the results we shall obtain are mainly concerned with the hyperbolic groups and are either easy or false for the Euclidean and spherical groups. For instance, we shall extend the following theorem of Howson **(7)** to discrete groups of motions: if S and T are finitely generated subgroups of a free group, then $S \cap T$ is also finitely generated. This theorem is trivial for the Euclidean and spherical groups, which contain no infinitely generated subgroups. We shall generalize the theorem of Karrass and Solitar **(8)** that if F is a free group and H is a finitely generated subgroup which contains a normal subgroup of F, then H is of finite index. This theorem is trivial for the spherical groups and is false for most of the Euclidean groups. For the above reasons we shall consider only the case of discrete hyperbolic groups. We shall usually omit "discrete hyperbolic." We mention the interesting result of Nielsen **(11)**, Bundgaard **(1)**, and Fox **(5)** that the above groups all contain subgroups of finite index with no elements of finite order.

2. Hyperbolic groups. Let D be the disk $\{z \mid |z| < 1\}$ in the complex plane, $\bar{D}$ its closure and E its boundary. D can be given a Riemannian metric so that it becomes the Poincaré model of the hyperbolic plane. The geodesics, which we shall call h-lines, are arcs of circles orthogonal to E. The isometries are the linear fractional transformations which preserve D. They are of the forms

Received June 29, 1959. This is in part taken from a doctoral dissertation presented to Yale University.

$$S(z) = \frac{az + \bar{b}}{bz + \bar{a}} \quad \text{and} \quad T(z) = \frac{c\bar{z} + \bar{d}}{d\bar{z} + \bar{c}},$$

where $a\bar{a} - b\bar{b} = c\bar{c} - d\bar{d} = 1$.

The transformation S is called a translation if it has two fixed points which are on E. This is equivalent to the condition $|a + \bar{a}| > 2$. A translation maps each circle through the fixed points onto itself. In particular, the h-line through the fixed points is invariant and is called the axis of the transformation. S is called a rotation if it has a fixed point in D, and a limit-rotation if it has a single fixed point on E. These conditions are respectively equivalent to $|a + \bar{a}| < 2$ and $|a + \bar{a}| = 2$. T is a reflection in an h-line if $d + \bar{d} = 0$. Otherwise T is a glide-reflection, that is, the product of a translation along an axis λ with a reflection in λ.

For each transformation S or T, there are a pair of h-lines λ and λ', called the isometric circles of the transformation (see **(4)**). S is the product of a reflection in λ and a reflection in the perpendicular bisector of the Euclidean line through the centres of the circles λ and λ'. If T is a glide-reflection, this product must be combined with a reflection in the h-line through the fixed points of T; if T is a reflection, $\lambda = \lambda'$ is the h-line of reflection. S is a translation, rotation or limit-rotation, according as λ and λ' do not intersect, do intersect, or are tangent. A discrete, hyperbolic group G has a canonical fundamental region, denoted R_G, which consists of the region in D outside of the isometric circles of all elements of G.

A subset M of $\bar{D}$ is called *h-convex* if with every two of its points it contains the h-line segment between them. For any subset M of $\bar{D}$, we denote by $[M]$, the h-convex closure of M, that is, the intersection of all h-convex subsets of $\bar{D}$ which contain M.

The set of limit points L_G of a group G is the intersection with E of the set of limit points of $\{g(z) | g \in G\}$, where z is any point in D. This set is independent of $z \in D$, because the transformations in G preserve hyperbolic distances, and these become arbitrarily small relative to Euclidean distance, as E is approached. L_G is a closed set, invariant under G. The *convex figure* of G is the set $K_G = [L_G] \cap D$. This is an h-convex set which is invariant under G.

For each limit-rotation $g \in G$, it is possible to find a limit-circle C_g so that:

(a) C_g is tangent to E at the fixed point of g,

(b) $C_g \subset K_G$,

(c) If g_1 and g_2 are limit-rotations such that $g_2 = fg_1f^{-1}$, where $f \in G$, then

$$C_{g_2} = fC_{g_1},$$

(d) If z_1 and z_2 are two points interior to C_g, u is the fixed point of g, and $z_2 = f(z_1)$, where $f \in G$, then f is either a limit-rotation with fixed point u, or f is a reflection in an h-line with one endpoint at u.

We shall denote by K^*_G the region obtained from K_G by deleting the interior of each C_g. K^*_G is neither unique nor h-convex, but it is invariant under G. We shall say that K^*_G *is compact mod* G, if there exists a disk $\Gamma = \{z \mid |z| < r < 1\}$ such that

$$K^*_G \subset G\Gamma = \bigcup_{g \in G} g\Gamma.$$

This is equivalent to the compactness of the surface obtained from K^*_G by identifying points congruent under G. Nielsen **(10; 12)** has proved that G *is finitely generated, if and only if* K^*_G *is compact* mod G.

It is not hard to see (by constructing the fundamental region) that every hyperbolic group $F(p; n_1, \ldots, n_d; r)$ is realized as a group without limit-rotations. In fact, according to Nielsen **(12)**, for any finitely generated, hyperbolic group G, there is a homeomorphism s of D, such that sGs^{-1} is a group of motions without limit-rotations. When G contains no limit-rotations, $K^*_G = K_G$.

3. The results. Coxeter **(2)** and Goldberg **(6)** have shown that every abelian subgroup of the modular group $F(0; 2, 3; 1)$ is cyclic. We shall prove the following stronger version of this for the discrete, orientation-preserving, hyperbolic groups.

THEOREM 1. *If* $F(p; n_1, n_2, \ldots, n_d; r)$ *is hyperbolic, then the centralizer of any element is cyclic. The possible finite orders are the divisors of* $n_1, n_2, \ldots, n_d$. *Any finite subgroup is a cyclic group, conjugate to a subgroup of* $\langle S_1 \rangle, \langle S_2 \rangle, \ldots,$ or $\langle S_d \rangle$.

Proof. It is well-known that two orientation-preserving linear fractional transformations commute if and only if they have the same fixed points. Therefore the centralizer of a rotation or limit-rotation is a group of rotations or limit-rotations with the same fixed point, and the centralizer of a translation is a group of translations with the same invariant axis. Each of these groups leaves a curve (or curves) invariant—a circle in D, for a group of rotations, a limit-circle for a group of limit-rotations, an h-line for a group of translations. Because the group is discrete, there must be an element which transforms a given point (on the invariant curve) the least distance in a fixed direction. This element generates the group, which is therefore cyclic.

The group $F(p; n_1, n_2, \ldots, n_d; r)$ has a fundamental region, which has among its vertices, the points $z_1, z_2, \ldots, z_d$ which are fixed points for $S_1, S_2, \ldots, S_d$ respectively. If z is a fixed point of a rotation S, there is an element f which maps z into one of the points z_k. Then fSf^{-1} is in the subgroup generated by S_k. Therefore the order of fSf^{-1}, which is the same as the order of S, divides n_k.

Let G be any finite subgroup of $F(p; n_1, \ldots, n_d; r)$, and let T_1 and T_2 be two elements (necessarily rotations) with fixed points t_1 and t_2 in D. We shall show that $t_1 = t_2$. Assuming otherwise, let λ_3 be the h-line through t_1 and t_2, and r_3 the reflection in λ_3. There are h-lines λ_2 and λ_1 through the points t_1

and t_2 respectively, such that if r_i is the reflection in λ_i, then $T_1 = r_2r_3$ and $T_2 = r_3r_1$. Therefore $T_1T_2 = r_2r_1$. If λ_1 and λ_2 diverge, r_2r_1 is a translation; if λ_1 and λ_2 are asymptotic (meet at a point on E), then r_2r_1 is a limit-rotation. Since these are transformations of infinite order, it follows that λ_1 and λ_2 must meet at a point t_3 in D, and $T_3 = r_1r_2$ is a rotation whose fixed point is t_3. The group $\langle r_1, r_2, r_3\rangle$ has the triangle $t_1t_2t_3$ as fundamental region. This group is infinite, since the images of $t_1t_2t_3$ under $\langle r_1, r_2, r_3\rangle$ cover D; since $\langle T_1, T_2, T_3\rangle$ is of index 2 in $\langle r_1, r_2, r_3\rangle$, the former subgroup is also infinite. We conclude that $t_1 = t_2$, and G is a cyclic group conjugate to a subgroup of $\langle S_1\rangle, \langle S_2\rangle, \ldots,$ or $\langle S_d\rangle$.

For hyperbolic groups which contain orientation-reversing transformations, the only exceptions are the following. The centralizer of a translation or glide-reflection can be a product of cyclic groups $C_\infty \times C_2$. The centralizer of a reflection can be the group $C_\infty \times C_2$ or $F(0; 2, 2; 1) \times C_2$. A finite subgroup can be a dihedral group.

THEOREM 2. *If S and T are finitely generated subgroups of a discrete group, then $S \cap T$ is also finitely generated.*

Proof. Let $H = S \cap T$ and let G be the finitely generated discrete group generated by S and T. As we remarked in § 2, we can suppose that G contains no limit-rotations. By Nielsen's theorem, there exist disks

$$\Gamma_S = \{z \mid |z| < r_s < 1\}, \qquad \Gamma_T = \{z \mid |z| < r_t < 1\}$$

such that $K_S \subset S\Gamma_S$ and $K_T \subset T\Gamma_T$. Let $r = \max(r_s, r_t)$ and $\Gamma = \{z \mid |z| < r\}$. Then

$$K_S \subset S\Gamma \quad \text{and} \quad K_T \subset T\Gamma.$$

Choose coset representatives $\{s_i\}$, $\{t_j\}$ so that

$$S = \bigcup_i Hs_i \qquad \text{and} \qquad T = \bigcup_j Ht_j.$$

Then

$$K_S \subset S\Gamma = H \bigcup_i s_i\Gamma,$$

$$K_T \subset T\Gamma = H \bigcup_j t_j\Gamma.$$

Also

$$K_H \subset K_S \cap K_T,$$

since

$$L_H \subset L_S \cap L_T.$$

We now show that $s_i\Gamma \cap K_T \neq \phi$ for only a finite number of representatives s_i. For any $h \in H$, $s_i\Gamma \cap ht_j\Gamma \neq \phi$ if and only if $\Gamma \cap s_i^{-1}ht_j\Gamma \neq \phi$. Now if $d(z_1, z_2)$ is the hyperbolic distance between the points z_1 and z_2 in D, and the

hyperbolic radius of Γ is ρ, then $\Gamma \cap g\Gamma \neq \phi$ if and only if $d(0, g(0)) < 2\rho$. But the discreteness of G implies that there are only a finite number of elements $g \in G$ with this last property. Therefore there are only a finite number of elements $g = s_i^{-1} h t_j$ with $\Gamma \cap g\Gamma \neq \phi$. Note that if

$$s_{i_1}^{-1} h_1 t_{j_1} = s_{i_2}^{-1} h_2 t_{j_2},$$

then

$$s_{i_2} s_{i_1}^{-1} h_1 = h_2 t_{j_2} t_{j_1}^{-1} \in S \cap T = H.$$

Therefore $s_{i_2} s_{i_1}^{-1}$ and $t_{j_2} t_{j_1}^{-1} \in H$, so

$$s_{i_1} = s_{i_2}, \qquad t_{j_1} = t_{j_2}$$

and $h_1 = h_2$. It follows that there are only a finite number of the s_i, t_j, h for which $s_i \Gamma \cap h t_j \Gamma \neq \phi$, and therefore only a finite number of the s_i for which $s_i \Gamma \cap K_T \neq \phi$.

Since $K_H \subset K_T$, there are only a finite number of the s_i, say $s_{i_1}, s_{i_2}, \ldots, s_{i_n}$, so that $s_i \Gamma \cap K_H \neq \phi$. Furthermore, the elements of H map K_H and K_S onto themselves and consequently $K_S - K_H$ onto itself. It follows that $s_i \Gamma \cap K_H \neq \phi$ if and only if $H s_i \Gamma \cap K_H \neq \phi$. Recalling that

$$K_H \subset H \bigcup_i s_i \Gamma,$$

we now obtain

$$K_H \subset H \bigcup_{k=1}^{n} s_{i_k} \Gamma.$$

Let Γ' be a disk with centre 0 and radius $r' < 1$, which is large enough to contain

$$\bigcup_{k=1}^{n} s_{i_k} \Gamma.$$

Then $K_H \subset H\Gamma'$, or K_H is compact mod H. Nielsen's theorem now implies that H is finitely generated.

THEOREM 3. *If H is a finitely generated subgroup of G and if $L_H = L_G$, then $[G: H]$ is finite.*

Proof. If $L_G = \phi$, then G must be finite. If L_G consists of a single point z, then the elements of G and H are limit-rotations whose fixed point is z, and possibly reflections in h-lines with one endpoint at z. It is easy to see that the index $[G: H]$ is finite in this case. If L_G contains more than one point, then K^*_G and K^*_H are non-empty sets. By Nielsen's theorem there is a disk $\Gamma = \{z \mid |z| < r < 1\}$ so that $K^*_H \subset H\Gamma$. Since G is discrete, there can be only a finite number of elements $g \in G$ so that $\Gamma \cap g\Gamma \neq \phi$. We shall show that every $g \in G$ is congruent mod H to one of these elements, which we denote by $g_1, g_2, \ldots, g_n$. Let $z \in \Gamma \cap K^*_H$ (we suppose that Γ is large enough

so that this intersection is not empty) and let $g \in G$. Since $K_G = K_H$, we have $K^*_G \subset K^*_H$. K^*_G is invariant under G, so $g(z) \in K^*_H$. Therefore there exists $h \in H$ so that $hg(z) \in \Gamma$. Thus $\Gamma \cap hg\Gamma \neq \phi$ (since $hg(z) \in \Gamma \cap hg\Gamma$) and $hg = g_k$ for some k. It follows that $[G: H]$ is finite.

The following is proved in **(4**, p. 43**)**.

LEMMA 1. *If S is a closed subset of E which contains more than one point, and S is invariant under a group G, then $S \supset L_G$.*

Definition. An N-chain of a group G is a sequence of subgroups $G_1, G_2, \ldots, G_n$ such that:

(a) $G_k \neq \{1\}$ $(k = 1, 2, \ldots, n)$,

(b) either G_k is a normal subgroup of G_{k+1}, or G_{k+1} is a normal subgroup of G_k.

We shall say that two subgroups H and K are N-equivalent if there is an N-chain $H = G_1, G_2, \ldots, G_n = K$. A subgroup which is N-equivalent to G will be called an N-subgroup.

We shall call a group *quasi-abelian* if it leaves invariant an h-line or a point in D. Such a group is either abelian or has an abelian subgroup of index 2. G is quasi-abelian if and only if L_G consists of 0, 1, or 2 points. The following Lemma shows than an N-equivalence class consists entirely of quasi-abelian groups if it contains one such group.

LEMMA 2. *If G and H are N-equivalent subgroups of a discrete group and G is not quasi-abelian, then $L_G = L_H$.*

Proof. Let the N-chain be $G = G_1, G_2, \ldots, G_n = H$. We proceed to prove by induction that

$$L_{G_k} = L_G \qquad (k = 1, 2, \ldots, n).$$

Clearly $L_{G_1} = L_G$. Assume $L_{G_k} = L_G$. If $G_k \subset G_{k+1}$, then $L_{G_k} \subset L_{G_{k+1}}$. On the other hand, L_{G_k} is invariant under G_{k+1}. For let $g \in G_{k+1}$ and $z_0 \in L_{G_k}$. There is a sequence $\{h_j\} \subset G_k$ so that for any $z \in D$,

$$\lim_{j \to \infty} h_j(z) = z_0.$$

Now $gh_jg^{-1} \in G_k$ and

$$\lim_{j \to \infty} h_j g^{-1}(z) = z_0,$$

so that

$$\lim_{j \to \infty} gh_jg^{-1}(z) = g(z_0).$$

Therefore $g(z_0) \in L_{G_k}$, and L_{G_k} is invariant under G_{k+1}. Since $L_{G_k} = L_G$ and G is not quasi-abelian, L_{G_k} contains more than 2 points. By Lemma 1,

$$L_{G_k} \supset L_{G_{k+1}}.$$

Thus

$$L_G = L_{G_k} = L_{G_{k+1}}.$$

It remains to consider the case where G_{k+1} is a normal subgroup of G_k. In this case $L_{G_{k+1}} \subset L_{G_k}$. Moreover, in the same manner as above we can show that $L_{G_{k+1}}$ is invariant under G_k.

We assert that $L_{G_{k+1}}$ contains more than one point. If $L_{G_{k+1}} = \phi$, then G_{k+1} is a finite group. G_{k+1} is either a group of rotations (and possibly reflections) with a common fixed point $z \in D$ or a reflection group of order 2. In the first case the point z must be invariant under all transformations in G_k. Then G_k is also a finite group, so that

$$L_G = L_{G_k} = \phi.$$

But this implies that G is quasi-abelian. In the second case, G_{k+1} consists of the identity and a reflection r in some h-line λ. The elements of G_k leave λ invariant. L_{G_k} is either empty or consists of the endpoints of λ. This is true also of L_G, so that G must be quasi-abelian. If $L_{G_{k+1}}$ contains only a single point z, then this point is invariant under G_k. G_k is a group of limit-rotations with limit-centre z (and possibly reflections in h-lines with one endpoint at z). Then

$$L_G = L_{G_k} = \{z\}$$

and it follows that G is quasi-abelian.

Lemma 1 now implies that $L_{G_{k+1}} \supset L_{G_k}$, so that

$$L_G = L_{G_k} = L_{G_{k+1}}.$$

It now follows that $L_H = L_{G_k} = L_G$.

The previous lemma and Theorem 3 imply the following.

THEOREM 4. *Let H be a finitely generated N-subgroup of a non-quasi-abelian group G. Then $[G:H]$ is finite.*

LEMMA 3. *If U and V are subnormal subgroups of a non-quasi-abelian group, then $U \cap V \neq \{1\}$.*

Proof. Let F_1, U_1, and V_1 be the orientation-preserving subgroups of index 2 in F, U, and V respectively. In the proof of Lemma 2, we saw that if an N-subgroup of F is a reflection group or order 2, then F is quasi-abelian. Therefore neither U nor V are reflection groups, so U_1 and V_1 are non-trivial, subnormal subgroups of F_1. There exist normal series

$$F_1 \supset F_2 \supset \ldots \supset F_n = U_1,$$

$$F_1 \supset F_2' \supset \ldots \supset F_n' = V_1,$$

where some of the F_k or some of the F_k' might coincide. We shall prove inductively that $F_k \cap F_k'$ is a non-trivial non-abelian group. If $F_2 \cap F_2' = \{1\}$,

then each element of F_2 commutes with each element of F_2' But two orientation-preserving transformations commute, if and only if they have the same fixed points. This implies that F_2 (and F_2') is a commutative group. The elements of F_2 must be rotations with a common fixed point $z_1 \in D$, limit-rotations with a common fixed point $z_2 \in E$, or translations with a common axis λ. Since F_2 is normal in F_1, the elements of F_1 must have the same invariant point or h-line. Therefore F_1 is abelian, and F is quasi-abelian. From this contradiction we conclude that $F_2 \cap F_2' \neq \{1\}$. Furthermore $F_2 \cap F_2'$ is not abelian, since this together with its normality in F_1 would imply that F_1 is abelian. Now suppose that $F_k \cap F_k' \neq \{1\}$ and is non-abelian. $F_{k+1} \cap F_k'$ and $F_k \cap F_{k+1}'$ are normal subgroups of $F_k \cap F_k'$. By the same argument as before, we conclude that $F_{k+1} \cap F_{k+1}' = (F_{k+1} \cap F_k') \cap (F_k \cap F_{k+1}') \neq \{1\}$ and is not abelian. It now follows that $U \cap V \neq \{1\}$.

THEOREM 5. *Let H and K be two non-quasi-abelian subgroups of a discrete group. Then H and K are N-equivalent, if and only if there is a non-trivial subgroup J which is simultaneously subnormal in H and K.*

Proof. The "if" part is obvious; we shall prove the "only if" part. There is an N-chain $H = G_1, G_2, \ldots, G_n = K$. The series

$$G_1 \supset G_1 \cap G_2 \supset G_1 \cap G_2 \cap G_3 \supset \ldots \bigcap_{k=1}^{n} G_k$$

is a normal series. We shall prove inductively that

$$\bigcap_{k=1}^{m} G_k$$

is a non-trivial, subnormal subgroup of G_m. This is certainly true for $m = 1$; assume that this is true for $m = p$. If G_p is a normal subgroup of G_{p+1}, then

$$\bigcap_{k=1}^{p+1} G_k = \bigcap_{k=1}^{p} G_k \neq \{1\}.$$

Since

$$\bigcap_{k=1}^{p+1} G_k$$

is subnormal in G_p, which is normal in G_{p+1}, it follows that

$$\bigcap_{k=1}^{p+1} G_k$$

is subnormal in G_{p+1}. Now suppose that G_{p+1} is a normal subgroup of G_p. G_p cannot be quasi-abelian. The conditions of Lemma 3 are fulfilled, with

$$F = G_p, \qquad U = \bigcap_{k=1}^{p} G_k, \qquad V = G_{p+1}.$$

Therefore

$$\bigcap_{k=1}^{p+1} G_k \neq \{1\}.$$

Since

$$\bigcap_{k=1}^{p} G_k$$

is subnormal in G_p,

$$\bigcap_{k=1}^{p} G_k \cap G_{p+1}$$

is subnormal in $G_p \cap G_{p+1} = G_{p+1}$. It now follows that the group

$$J = \bigcap_{k=1}^{n} G_k$$

is a non-trivial, subnormal subgroup of H and K.

This theorem implies that if G is not quasi-abelian, then a subgroup H is an N-subgroup if and only if it contains a subnormal subgroup of G.

THEOREM 6. *Let H be a finitely generated non-quasi-abelian subgroup of G. Then there is a subgroup G_H of G such that*

(a) *G_H is N-equivalent to H,*

(b) *if $K \subset G$ and K is N-equivalent to H, then $K \subset G_H$,*

(c) *$[G_H : H]$ is finite.*

Proof. Let $G_H = \{g \mid g \in G, gL_H = L_H\}$. Since

$$H \subset G_H, L_H \subset L_{G_H}.$$

Lemma 1 implies that $L_H \supset L_{G_H}$, so that $L_H = L_{G_H}$. Theorem 3 now implies that $[G_H : H]$ is finite. From this it follows that H has a finite number of conjugate subgroups in G_H. The intersection of these conjugate subgroups is a normal subgroup F of finite index in G_H. Since G_H is infinite, F is non-trivial. Therefore the sequence, G_H, F, H, is an N-chain, and G_H is N-equivalent to H. If K is N-equivalent to H, Lemma 2 implies that K leaves L_H invariant, so that $K \subset G_H$.

THEOREM 7. *Let H and K be finitely generated non-quasi-abelian subgroups of a discrete group. Then the following statements are equivalent:*

(a) *H and K are N-equivalent;*

(b) *there is a group J which is simultaneously normal and of finite index in H and K;*

(c) *$L_H = L_K$.*

Proof. If (a) *is true,* then $G_H = G_K$. (These are the groups introduced in Theorem 6.) Since H and K are of finite index in G_H, this is also true of $H \cap K$. Therefore $H \cap K$ contains a nontrivial subgroup J which is normal and of finite index in G_H. J is also normal and of finite index in H and K. This shows that (a) implies (b).

If (b) is true, then H and K are N-equivalent. Therefore $L_H = L_K$. This shows that (b) implies (c).

Now suppose (c) is true, Then $G_H = G_K$. It follows that H and K are both N-equivalent to $G_H = G_K$, and hence to each other.

It would be interesting to determine whether there are algebraic conditions equivalent to the condition $L_H = L_K$, when H or K is infinitely generated.

The following is proved in **(9**, p. 76**)**.

LEMMA 4. *Let U and V be two groups such that the isometric circles of U are contained in R_V and the isometric circles of V are contained in R_U. Then the group generated by U and V is the free product $U * V$, and $R_{U*V} = R_U \cap R_V$.*

THEOREM 8. *Let H be a finitely generated subgroup of a finitely generated non-quasi-abelian group G. Then $[G:H]$ is finite if and only if H is contained in no infinitely generated subgroup of G.*

Proof. If H is of finite index, then any larger group must also be of finite index, and so it is finitely generated.

Now suppose H is of infinite index. We shall find a subgroup of G which contains H and is infinitely generated. We may assume that G contains no limit-rotations. Then $\bar{R}_H$, which cannot be contained in D, contains intervals on E. We first show that one of these intervals contains points of L_G in its interior.

If $L_H = L_G$, then by Theorem 3 $[G:H]$ is finite. Thus there is a point $z_0 \in L_G - L_H$. Let $z \in R_H$; there is a sequence $\{g_n\} \subset G$ such that

$$\lim_{n\to\infty} g_n(z) = z_0.$$

Since R_H is a fundamental region for H, there is $h_n \in H$ so that $h_n g_n(z) \in R_H$. The sequence $\{h_n g_n(z)\}$ has a subsequence which converges to a point $z_1 \in \bar{R}_H \cap E$. The point z_1 is a limit point of G and belongs to an interval I_1 of $\bar{R}_H \cap E$. z_1 might possibly be an endpoint of I_1. Since G is not quasi-abelian, L_G consists of more than two points, and hence it is a perfect subset of E (see **(4**, p. 68**)**). Thus there is a sequence $\{x_n\}$ in L_G which converges to z_1. Suppose this sequence is outside I_1. z_1 is the endpoint of an isometric circle of an element $h \in H$. The transformation h maps I_1 outside $\bar{R}_H \cap E$, and maps a neighbouring interval, containing almost all of the sequence $\{x_n\}$, onto an interval I of $\bar{R}_H \cap E$. Therefore I contains points of L_G in its interior.

As is shown in **(10)**, the fixed points of the translations of G are dense in L_G in the following sense. If $x, x' \in L_G$ and J and J' are intervals of E which contain x and x' respectively, then there is a translation $g \in G$, with a fixed point in each interval. A sufficiently high power g^n has isometric circles λ and λ' which intersect E inside J and J' respectively. Since the interval $I \subset \bar{R}_H \cap E$ contains points in L_G, it contains an infinite sequence of such points, which we denote by $\{y_1, y_1', y_2, y_2', \ldots,\}$. Let I_k, I_k' be mutually disjoint subintervals of I which contain y_k and y_k' respectively. There is a

translation $g_k \in G$ whose isometric circles λ_k and λ_k' intersect E inside I_k and I_k' respectively. By Lemma 4, the group generated by $\{g_1, g_2, \ldots,\}$ is a free group F of infinite rank, whose fundamental region R_F is the region in D outside of all λ_k and λ_k'. Lemma 4 now implies that the group K generated by H and F is the free product $H * F$. Thus K is an infinitely generated group containing H.

THEOREM 9. *Let H be a finitely generated subgroup of a non-quasi-abelian group G. If H has a non-trivial intersection with every non-cyclic subgroup of G, then $[G:H]$ is finite.*

Proof. We shall show that $L_H = L_G$. Since G is not quasi-abelian, L_G is a perfect subset of E. Let $z \in L_G$, and let I be an open interval of E which contains z. I contains infinitely many points of L_G. Choose four of them z_1, z_1', z_2, z_2'. Let I_1, I_1', I_2, I_2' be non-intersecting subintervals of I, which contain z_1, z_1', z_2, z_2' respectively. As in the proof of Theorem 8, there are translations g_1 and $g_2 \in G$, such that the isometric circles λ_1 and λ_1' of g_1 intersect E inside I_1 and I_1' respectively, and the isometric circles λ_2 and λ_2' intersect E inside I_2 and I_2' respectively. The group K, generated by g_1 and g_2, is a free group of rank 2. By hypothesis, the intersection $H \cap K$ is non-trivial. It follows that H has an element whose fixed points are in I. Since this is true for any interval I containing z, it follows that $z \in L_H$ and $L_H = L_G$. Theorem 3 now implies the required result.

COROLLARY. *Let H and K be finitely generated non-quasi-abelian subgroups of a discrete group. If H has a non-trivial intersection with every non-cyclic subgroup of K, then $[K: H \cap K]$ is finite.*

Proof. By Theorem 2, $H \cap K$ is finitely generated. The Corollary now follows from Theorem 9.

References

1. S. Bundgaard and J. Nielsen, *On normal subgroups with finite index in F-groups*, Mat. Tids. B (1951), 56–58.
2. H. S. M. Coxeter, *On subgroups of the modular group*, J. de Math. Pures et App. (1958), 317–319.
3. H. S. M. Coxeter and W. O. J. Moser, *Generators and relations for discrete groups*, Ergeb. der Math. (1957).
4. L. Ford, *Automorphic functions* (New York, 1951).
5. R. Fox, *On Fenchel's conjecture about F-groups*, Mat. Tids. B (1952), 61–65.
6. K. Goldberg, *Unimodular matrices or order 2 that commute*, J. Washington Acad. Sci., *46* (1956), 337–338.
7. A. G. Howson, *On the intersection of finitely generated free groups*, J. London Math. Soc., *29* (1954), 428–434.
8. A. Karrass and D. Solitar, *Note on a theorem of Schreier*, Proc. Amer. Math. Soc., *8* (1957), 696–697.
9. W. Magnus, *Discrete groups* (New York University Notes, 1952).
10. J. Nielsen, *Ueber Gruppen linearer Transformationen*, Mitteilungen der Math. Ges. in Hamburg Band VIII (1940), 82–104.
11. ——— *Kommutatorgruppen for det frie product af cykliske grupper*, Mat. Tids. B (1948), 49–56.
12. ——— *Nogle grundlaeggende begreber vedrørende diskontinuerte grupper af lineaere substitutioner i en kompleks variabel*, Den II^te Skandinaviske Matematikerkongress i Trondheim (1949), 61–70.

Brown University

LIMITS OF LATTICES IN A COMPACTLY GENERATED GROUP

A. M. MACBEATH AND S. ŚWIERCZKOWSKI

1. Introduction. Let G be a locally compact and σ-compact[1] topological group and let H be a discrete subgroup of G.[2] We shall use G/H to denote the space of right cosets Hx of H with the usual topology (cf. **(8**, pp. 26–28**)**). Let μ be the left Haar measure in G. μ induces a measure in the space G/H;[3] this measure will, without ambiguity in this paper, also be denoted by μ. If $\mu(G/H)$ is finite, the group H is called a *lattice*. If the space G/H is compact, then H is certainly a lattice and is called a *bounded lattice*. These terms are an extension of the usage of the Geometry of Numbers, where G is the real n-dimensional vector space R^n. In this case any lattice is generated by n linearly independent vectors, all lattices are bounded, and the whole family of lattices is permuted transitively by the automorphisms of G (which are the non-singular linear transformations). The constant $\mu(G/H)$ is called the determinant of H in this case. The family of all lattices in Euclidean space forms a locally compact topological space. In **(7)** Mahler proved the following

SELECTION THEOREM. *Let $\{H_n\}$ be a sequence of lattices in R^n with the following properties*

(i) *There is a neighbourhood V of the zero-vector e such that, for all n, $H_n \cap V = \{e\}$,*

(ii) *$\mu(G/H_n)$ is bounded above.*

Then there exists a subsequence $\{H_{n'}\}$ of $\{H_n\}$ which converges to a lattice H.

Let now G, H, and μ be as in the beginning. Mahler's theorem suggests two definitions. [Notation: e is the unity of G, $\mathbf{N}$ the class of open sets containing e; $\bar{K}$, fr$K$ are closure and boundary of K; $\cup$, $-$, $\cap$ denote the set union, difference, intersection. We use $\varphi: G \to G/H$ for the natural mapping $\varphi(x) = Hx$.]

DEFINITION 1. A sequence $\{H_n\}$ of subgroups of G is called *uniformly discrete* if $H_n \cap V = \{e\}$ for a certain $V \in \mathbf{N}$ and all n.

DEFINITION 2. A sequence $\{H_n\}$ of subgroups of G *converges to a subgroup H* if, given any compact set C and any $V \in \mathbf{N}$,

$$H \cap C \subset H_n V \quad \text{and} \quad H_n \cap C \subset HV$$

holds for all but a finite number of n.

Received April 27, 1959.

[1] That is, G is a countable union of compact sets.

[2] As is well known, this implies that H is countable.

[3] We give a precise definition of the induced measure in § 3.

Chabauty **(2)** has generalized Mahler's theorem by showing that a uniformly discrete sequence $\{H_n\}$ of subgroups of G has a subsequence converging to a discrete subgroup H and moreover

$$\mu(G/H) \leqslant \liminf \mu(G/H_n) \tag{1}$$

so that H is a lattice if all the H_n are and $\mu(G/H_n)$ is bounded.

In the classical case $G = R^n$, it is of course easy to show that

$$\mu(G/H) = \lim \mu(G/H_n) \tag{2}$$

and Chabauty has shown that in certain circumstances this is true also for topological groups G. In this paper we make a further contribution to this problem by proving that, if H is a bounded lattice, then a necessary and sufficient condition for (2) to hold is that G should be compactly generated[4] or that H should be finitely generated. We shall give an example due to M. Kneser showing that the boundedness of H is essential. Thus it might seem better to consider bounded lattices only, particularly since in Geometry of Numbers all lattices are bounded. Unfortunately however, a lattice which is a limit of bounded lattices need not be bounded. In § 6 we shall give an example of such a lattice where G is a homomorphic image of the group of 2 by 2 matrices with determinant unity.

2. Fundamental domain. As in **(5)** and **(10)** a Borel set P will be called a packing if $P \cap hP = \phi$ for $e \neq h \in H$ and a Borel set C will be called a *covering* if $HC = G$. F is called a *fundamental domain* if it is both a packing and a covering. In cases of ambiguity we may refer to an H-packing, H-covering, or H-fundamental domain.

In this section we show, extending a result of Chabauty **(1)**, and Siegel **(9)**, that there is a fundamental domain F with $\mu(\mathrm{fr}F) = 0$, and also that if G/H is compact then there is such a fundamental domain with compact closure $\bar{F}$.

We shall overlap in places with Chabauty's results. We start with a lemma which shows that Chabauty's axiom (M) is always satisfied.

LEMMA 2.1. *If C is compact and U is open, $C \subset U$, then there is a Baire measurable open set V such that*

$$C \subset V \subset U, \quad \mu(\mathrm{fr}V) = 0.$$

In particular, taking $C = \{e\}$, there is a fundamental system of neighbourhoods of the identity each of which has a frontier of measure 0.

Proof. Since the measure μ is regular, and the measure of any compact set is finite **(4, §§ 64 and 52)**, we may assume, on replacing U by an open subset, if necessary, that $\mu(U) < \infty$. Since the group G is a completely regular space **(8, p. 29)**, a continuous function $f(x)$ exists such that $f(x) = 0$

[4] That is, have a compact set of generators.

for $x \in C$, $f(x) = 1$ for $x \notin U$. Let $E(r) = \{x: f(x) < r\}$. The function $\mu(E(r))$ is a monotonic function of the real variable r, and therefore has at most countably many discontinuities. Let r_0 be a value at which it is continuous. Then

$$\overline{E(r_0)} = \bigcap \{W: W \text{ open}, W \supset E(r_0)\} \subset \bigcap_{r>r_0} E(r).$$

Hence

$$\mu(E(r_0)) \leqslant \mu(\overline{E(r_0)}) \leqslant \lim_{r \to r_0} \mu(E(r)) = \mu(E(r_0)).$$

This completes the proof, since $V = E(r_0)$ is a Baire set.

The following two lemmas are easily verified:

LEMMA 2.2. *If A, B are packings and $C = (A - HB) \cup B$, then C is a packing and $HC = HA \cup HB$.*

LEMMA 2.3.

$$B \cap \mathrm{fr}A \subset \mathrm{fr}(A \cap B) \cup \mathrm{fr}B.$$

We begin now the construction of a fundamental domain. Our final result will be as follows.

THEOREM 1. *There is a fundamental domain F such that*
(i) $\mu(\mathrm{fr}F) = 0$;
(ii) *If G/H is compact, there exists a fundamental domain F satisfying* (i) *such that also $\bar{F}$ is compact.*

The proof of this Theorem is closely modelled on that of Siegel **(9)**.

Proof. Since G is locally compact and H is discrete, we can, by Lemma 2.1, choose $V \in \mathbf{N}$ so that $\mu(\mathrm{fr}V) = 0$, $\bar{V}$ is compact, and V is an H-packing. Since G is σ-compact, $G \subset \bigcup Vx_i$ for some sequence $\{x_i\} \subset G$. Define $F_1 = Vx_1$, $F_n = Vx_n - H(Vx_1 \cup \ldots \cup Vx_{n-1})$. Let $F = \bigcup F_n$. Then clearly F is an H-packing, since F_n is and since $F_m \cap hF_n = \phi$. Also $HF = G$, for if $g \in G$, there is a least integer n such that $g \in HVx_n$ and then $g \in HF_n \subset HF$. Thus F is a fundamental domain.

To show that $\mu(\mathrm{fr}F) = 0$, set $C_n = Vx_1 \cup \ldots \cup Vx_{n-1}$. Then $\mathrm{fr}C_n \subset \mathrm{fr}Vx_1 \cup \ldots \cup \mathrm{fr}Vx_{n-1}$, so $\mu(\mathrm{fr}C_n) = 0$. Also

$$F_n = Vx_n - HC_n = Vx_n - \bigcup_{h \in H} (hC_n \cap Vx_n).$$

If empty terms are dropped from the last union, only those h remain for which $h \in Vx_nC_n^{-1}$. Since $Vx_nC_n^{-1}$ is a bounded set, the number of h is finite, say $h_1, \ldots h_r$, and we have

$$F_n = Vx_n - (h_1C_n \cup \ldots \cup h_rC_n)$$

$$\mu(\mathrm{fr}F_n) \leqslant \mu(\mathrm{fr}Vx_n) + \sum_{i=1}^{r} \mu(\mathrm{fr}h_iC_n) = 0.$$

By Lemma 2.3, $Vx_n \cap \mathrm{fr}F \subset \mathrm{fr}(Vx_n \cap F) \cup \mathrm{fr}Vx_n = \mathrm{fr}F_n \cap \mathrm{fr}Vx_n$. Thus $\mu(\mathrm{fr}F) \leqslant \sum\mu(Vx_n \cap \mathrm{fr}F) = 0$.

In the case when G/H is compact, G/H can be covered by a finite union $\varphi(Vx_1) \cup \ldots \cup \varphi(Vx_n)$, so $F = F_1 \cup \ldots \cup F_n$ will be a fundamental domain. Since F is then contained in the bounded set C_{n+1}, it is itself bounded. This completes the proof of Theorem 1.

We conclude this section with a slightly more precise form of the statement of Theorem 1. This is required for a later application.

LEMMA 2.4. *If S is any covering, then there is a fundamental domain contained in S.*

Proof. Let F be any fundamental domain. We have $F \subset HS$. Thus F is a union of h-translates of subsets of S and therefore F is also a disjoint union of h-translates of subsets of S, say $F = h_1S_1 \cup h_2S_2 \cup \ldots$. It is obvious that $F_0 = S_1 \cup S_2 \cup \ldots$ is a fundamental domain contained in S.

3. The induced measure in G/H. Since we regard the group H as a group of permutations acting on G by left translation, it follows that each H-orbit is a *right* coset Hx. This is why we use G/H for the space of right cosets, instead of the more usual homogeneous space of left cosets. On the space G/H the group G acts transitively by *right* translation. If $\Delta(x)$ is the real-valued function defined on G by the relation $\mu(Ex) = \Delta(x) \cdot \mu(E)$, then it follows from the criterion in **(11**, p. 45**)** that there is a measure $\tilde{\mu}$ on Borel subsets $\tilde{E}$ of G/H such that $\tilde{\mu}(\tilde{E}x) = \tilde{\mu}(\tilde{E}) \cdot \Delta(x)$. For our purposes it is more convenient to define μ directly from the natural mapping $\varphi: G \to G/H$, $(\varphi(x) = Hx)$, as follows: If F is any fundamental domain, define

$$\tilde{\mu}(\tilde{E}) = \mu(\varphi^{-1}(E) \cap F).$$

It follows from **(5**, Theorem 1, Corollary**)**, applied to the measure space $\varphi^{-1}(E)$ and the group H of transformations of this space, that this expression does not depend on the particular fundamental domain chosen. We shall, for $S \subset G$, use S/H to denote $\varphi(S)$ and we shall write μ for $\tilde{\mu}$.

We conclude this section with three lemmas which will be useful later.

Before stating the first lemma, we note that if G_1 is any open subgroup of G, the same measure μ, but with its domain of definition restricted to G_1 will serve as a Haar measure on G_1.

LEMMA 3.1. *If G_1 is an open subgroup of G and $H_1 = H_1 \cap H$, then $\mu(G_1/H_1) = \mu(G_1/H)$.*

Proof. Let F_1 be a fundamental domain for H_1 in G_1. Then $hF_1 \cap F_1 \neq \phi$, $h \in H$, implies $h \in F_1F_1^{-1} \subset G_1$, so $h \in H_1$ and $h = e$. Thus F_1 is an H-packing. If F is a fundamental domain for H in G, then, so is $F^* = F_1 \cup (F - HF_1)$, by Lemma 2.2. By our definition of induced measure,

$$\mu(G_1/H) = \mu(F^* \cap G_1) = \mu(F_1) = \mu(G_1/H_1).$$

LEMMA 3.2. *If $H_1 \subset H$ and $H:H_1$ denotes the index of H_1 in H, then we have*

$$\mu(G/H_1) = (H:H_1)\mu(G/H).$$

Proof. Let F be an H-fundamental domain and let X be a complete system of representatives of left cosets of H_1 in H. One checks that XF is an H_1-fundamental domain and our result follows then since $\overline{\overline{X}} = H:H_1$.

LEMMA 3.3. *If $H \subset G_1$, where G_1 is an open subgroup of G, then*

$$\mu(G/H) = (G:G_1)\mu(G_1/H).$$

Proof. If G_1 is not unimodular, neither is G and $\mu(G/H) = \mu(G_1/H) = \infty$ (9, Lemma 5). Suppose next that G_1 is unimodular, but not G. Then, since G_1 is open, μ is also the Haar measure for G_1 and we have $\Delta(x) = 1$ for $x \in G_1$. However, if $\Delta(x) \neq 1$, where $x \in G$, then $\Delta(x^n) \neq 1$ for each natural n. All the elements x^n must then belong to different left cosets of G_1 and hence $G:G_1 = \infty$. Again both sides are infinite.

The remaining case to consider is when G is unimodular. Then, if F is an H-fundamental domain for G_1 and X is a complete system of representatives of right cosets of G_1, we verify that FX is an H-fundamental domain for G. Since G is unimodular our result follows from $\overline{\overline{X}} = G:G_1$.

LEMMA 3.4. *If G_1 is an open subgroup of G, then*

$$\mu(G/H) \leqslant (G:G_1)\mu(G_1/H).$$

Proof. Let $H_1 = G_1 \cap H$. By Lemmas 3.1, 3.2, and 3.3 we have

$$\mu(G/H_1) = (G:G_1)\mu(G_1/H_1) = (G:G_1)\mu(G_1/H),$$
$$\mu(G/H_1) = (H:H_1)\mu(G/H) \geqslant \mu(G/H).$$

This proves the lemma.

LEMMA 3.5. *If K is an open subgroup of G and HK is also a subgroup, then* $\mu(HK/H) = \mu(K/K \cap H)$.

Proof. Let F be a $(K \cap H)$-fundamental domain for the group K. One checks that F is an H-fundamental domain for HK.

4. Limits of discrete subgroups. In this section we assume G/H compact. We consider the following two closely related questions:

I. In what groups G does the relation $\lim H_n = H$ imply $\lim \mu(G/H_n) = \mu(G/H)$ for any uniformly discrete sequence of subgroups $\{H_n\}$?

II. Under what circumstances does $\lim H_n = H$ imply $\lim \mu(G/H_n) = \mu(G/H)$ if $\{H_n\}$ is restricted to be a uniformly discrete sequence of *lattices*?

Our answer to I is complete, given by the theorem below. As to question II we give a little extra information in Theorem 3. Another kind of answer was found by Chabauty and we present in § 5 an alternative proof of his result (our Theorem 4).

THEOREM 2. *The following four statements are equivalent:*

(i) *G is compactly generated.*

(ii) *H is finitely generated.*

(iii) *If $\{H_n\}$ is a sequence of discrete subgroups, $\lim H_n = H$, then*

$$\limsup \mu(G/H_n) \leqslant \mu(G/H).$$

(iv) *If $\{H_n\}$ is a uniformly discrete sequence of subgroups, $\lim H_n = H$, then*

$$\lim \mu(G/H_n) = \mu(G/H).$$

Proof. We have proved in a recent paper that (i) implies (ii) (see **6**). Suppose (ii) holds. It follows from Theorem 1 (ii) that there exists an H-fundamental domain F with compact closure $\bar{F}$. If Γ is the finite set of generators of H, then the compact set $\Gamma \cup \bar{F}$ is obviously a set of generators of G. Hence (ii) implies (i) and so (i) and (ii) are equivalent. By Chabauty's inequality (1), (iii) implies (iv). Thus it remains to prove that (iv) implies (ii) and that (i) implies (iii).

Proof that (iv) *implies* (ii). Suppose that (ii) is false. Then H being countable let its elements be enumerated $h_1, h_2, \ldots$, and let H_n be the subgroup generated by the elements $h_1, \ldots, h_n$. If C is a compact set, $C \cap H$ is finite and if n_0 is the largest value of r for which h_r lies in C, we have $C \cap H = C \cap H_n$ for $n > n_0$. Thus $\lim H_n = H$. However, the index $H:H_n$ is infinite, otherwise H would have a finite system of generators given by $h_1, \ldots, h_n$ together with a complete system of representatives of the H_n-cosets. It follows from Lemma 3.2 that $\mu(G/H_n) = \infty$ for all n. But $\mu(G/H) < \infty$, so (iv) is false.

Proof that (i) *implies* (iii). Let F be an H-fundamental domain with compact closure $\bar{F}$, such that $\mu(\mathrm{fr}F) = 0$. We have $\mu(G/H) = \mu(F) = \mu(\bar{F})$. Let $\epsilon > 0$. We have to show that, for sufficiently large n, $\mu(G/H_n) < \mu(G/H) + \epsilon$. Choose $V \in \mathbf{N}$, $\bar{V}$ compact, so that

$$\mu(VF) < \mu(F) + \epsilon. \tag{3}$$

Let D be a compact system of generators of G. Replacing D by $D \cup D^{-1}$, if necessary, we may assume that

$$\bigcup_1^\infty D^k = G. \tag{4}$$

The set $\overline{VFDF^{-1}}$ is compact, so there is a finite number of elements $h_1, \ldots, h_r$ of H in it. We have $VFD \subset HF$; but $hF \cap VFD = \phi$ unless $h \in VFDF^{-1}$, that is, unless h is one of the elements $h_1, \ldots, h_r$. It follows that

$$VFD \subset h_1F \cup \ldots \cup h_rF. \tag{5}$$

Since $\lim H_n = H$, there is a number n_0, such that, for $n > n_0$, H_nV contains each of the elements $h_1, \ldots, h_r$, and hence from (5), $VFD \subset H_nVF$. But H_n is a subgroup, $H_n = H_n^k$ for each integer k, and thus

$$VFD^k \subset H_nVFD^{k-1} \subset \ldots \subset H_n^{k-1}VFD \subset H_n^kVF = H_nVF.$$

Thus $G = H_nVF$ by (4). Hence VF is an H_n-covering and by the theorem on packings and coverings in **(5)** it follows from (3) that

$$\mu(G/H_n) \leqslant \mu(VF) \leqslant \mu(G/H) + \epsilon.$$

To state our next theorem briefly, it is convenient to have another definition. A pair (G, H) consisting of a locally compact σ-compact group G and a discrete subgroup H with G/H compact will be called a *tractable pair* if the following condition holds. Given any uniformly discrete sequence $\{H_n\}$ of lattices in G such that $\lim H_n = H$, then $\lim \mu(G/H_n) = \mu(G/H)$.

THEOREM 3. *If G contains an open compactly generated subgroup K such that for $h \in H$*

$$hKh^{-1} = K \tag{6}$$

then (G, H) is tractable if and only if (H, H) is tractable.

Proof. It is quite clear that if (H, H) is not tractable, then (G, H) is not tractable. For there will be a sequence $\{Q_n\}$ of subgroups of H of finite index such that $\lim Q_n = H$, but $H: Q_n > 1$ for infinitely many n. By Lemma 3.2, $\mu(G/Q_n) = (H: Q_n)\mu(G/H) \geqslant 2\mu(G/H)$ for infinitely many n. Thus (G, H) is not tractable.

We now assume therefore, that (H, H) is tractable and our aim is to prove that (G, H) is tractable. We shall show that if $\{H_n\}$ is a sequence of lattices in G and $\lim H_n = H$, then

$$\limsup \mu(G/H_n) \leqslant \mu(G/H). \tag{7}$$

Hence for a uniformly discrete sequence $\{H_n\}$ of lattices we have by (1), $\lim \mu(G/H_n) = \mu(G/H)$, that is, (G, H) is tractable.

Since the topology in H is discrete, our assumption that (H, H) is tractable means that, if $\{Q_n\}$ is a sequence of subgroups of H with the following properties:

(8) (i) $$H = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} Q_n,$$

(ii) $$H: Q_n < \infty,$$

then there is a number n_0 such that $H = Q_n$ for $n > n_0$.

Suppose now that $\{H_n\}$ is a sequence of lattices in G such that $\lim H_n = H$. To show (7) we shall associate with the sequence $\{H_n\}$ a sequence $\{Q_n\}$ of subgroups of H which satisfies the conditions (8). We observe first that, by (6), HK and

$$M_n = H_n \cap HK, \qquad P_n = M_nK, \qquad Q_n = H_nK \cap H$$

are subgroups of G and moreover P_n is open.

LEMMA 4.1. $H: Q_n = HK: P_n$.

Proof. One checks easily that any complete system of representatives of left cosets of Q_n in H is also a complete system of representatives of left cosets of P_n in HK.

LEMMA 4.2. *For $n > n_0$, $Q_n = H$, $P_n = HK$.*

Proof. By Lemma 4.1 it is enough to show that $Q_n = H$. Since (H, H) is tractable this follows if we show that conditions (8) are satisfied. To prove that Q_n has finite index, we note that, by Lemmas 3.1 and 3.3,

$$(HK: P_n)\mu(P_n/M_n) = \mu(HK/M_n) = \mu(HK/H_n) \leqslant \mu(G/H_n) < \infty.$$

Now P_n is for sufficiently large n a non-empty open set, so $\mu(P_n/M_n) > 0$, and by Lemma 4.1, $H_n: Q_n = HK: P_n < \infty$.

To show that (8) (i) holds we have to show that if $h \in H$, then, for sufficiently large n, $h \in Q_n$. To see this we note that $K \in \mathbf{N}$, so for sufficiently large n, $hK \cap H_n \neq \phi$, that is, $h \in H_nK$. This proves our lemma.

We are now in a position to prove (7). By Theorem 2, since K is compactly generated

$$\limsup \mu(K/K \cap H_n) \leqslant \mu(K/K \cap H). \tag{9}$$

From Lemma 3.5, we have $\mu(HK/H) = \mu(K/K \cap H)$. If, in Lemma 3.5 we replace H by M_n so that HK is replaced by P_n, we find that $\mu(P_n/M_n) = \mu(K/K \cap H_n)$. From Lemma 3.1, we have $\mu(P_n/H_n) = \mu(P_n/M_n)$ since $P_n \cap H_n = M_n$. Hence $\mu(P_n/H_n) = \mu(K/K \cap H_n)$ and substituting in (9) we derive

$$\limsup \mu(P_n/H_n) \leqslant \mu(HK/H). \tag{10}$$

For sufficiently large n we have, by Lemma 4.2,

$$\mu(HK/H_n) = \mu(P_n/H_n). \tag{11}$$

Using (10), (11), and Lemmas 3.3 and 3.4,

$$\begin{aligned}\mu(G/H) = (G: HK)\mu(HK/H) &\geqslant (G: HK) \limsup \mu(P_n/H_n) \\ &= (G: HK) \limsup \mu(HK/H_n) \geqslant \limsup \mu(G/H_n).\end{aligned}$$

This completes our proof.

5. A result of Chabauty. We shall give now an alternative proof of a theorem of Chabauty **(1)** which combined with (1) yields another kind of answer to our question II.

THEOREM 4. *If $\{H_n\}$ is a sequence of lattices, $\lim H_n = H$ and there exists a set S of finite measure which is an H_n-covering for each n, then*

$$\limsup \mu(G/H_n) \leqslant \mu(G/H).$$

Proof. Let F, F_n denote the H and H_n-fundamental domains so that

$$\mu(G/H_n) = \mu(F_n), \qquad \mu(G/H) = \mu(F).$$

By Lemma 2.4 we may assume $F_n \subset S$. From $S \subset HF$ follows that we can cover S, except for a set of arbitrarily small measure, by a finite union $h_1F \cup \ldots \cup h_mF$, $h_i \in H$. Since $H = \lim H_n$ it follows that these sets in turn can be approximated by unions

$$h_1^{(n)}F \cup \ldots \cup h_m^{(n)}F, \qquad \text{where} \qquad h_i^{(n)} \in H_n.$$

Therefore, for sufficiently large n, an arbitrarily small part of S remains uncovered by H_nF. Hence, by $F_n \subset S$, we have $\lim [\mu(F_n) - \mu(F_n \cap H_nF)] = 0$. Since

$$\mu(F_n \cap H_nF) = \mu\Big(\bigcup_{H_n} (F_n \cap hF)\Big) \leqslant \sum_{H_n} \mu(F_n \cap hF) = \sum_{H_n} \mu(h^{-1}F_n \cap F)$$

$$= \mu(H_nF_n \cap F) = \mu(F)$$

the theorem follows.

6. Examples. In this section we give three examples illustrating different possible properties of convergent sequences of discrete subgroups.

Example 1. It follows from Theorem 2 that, if G is compactly generated, G/H_n compact and $\lim H_n = H$, then $\limsup \mu(G/H_n) \leqslant \mu(G/H)$. To show that this need not be true if G is not compactly generated, take $G = H = G_1 \times G_2 \times \ldots \times G_n \times \ldots$, the weak direct product of a countable family of cyclic groups of order 2, with the discrete topology. Define H_n to be the set of all $g = (g_1, g_2, \ldots, g_n, \ldots,) \in G$ with $g_n = e$. Then $\mu(G/H_n) = 2$, $\lim H_n = H$, $\mu(G/H) = 1$.

Example 2. In this example G is a connected Lie group, and G/H_n is compact for each n, but G/H is not compact. Let G be the group of all linear transformations

$$w = \frac{az + b}{cz + d},$$

where w, z are complex variables, a, b, c, d are real and $ad - bc > 0$. In addition to G we consider the set G_1 of inversions, that is, transformations of the form

$$w = \frac{(a\bar{z} + b)}{(c\bar{z} + d)},$$

where $\bar{z}$ is the complex conjugate, and a, b, c, d are real with $ad - bc < 0$. The set $G \cup G_1$ is a group of transformations of the upper half-plane $\Re z > 0$ on itself, and G is a normal subgroup of index 2. The topology is the natural one obtained from the variables a, b, c, d.

Let P be the point $i = \sqrt{-1}$, and let $Q = ki (1 < k < \sqrt{3})$ be a variable point on the imaginary axis. Let $C(Q)$ be the circle through Q with centre on the positive real axis and cutting the imaginary axis at an angle $\frac{1}{3}\pi$. Let $C(Q)$ cut $|z| = 1$ in R and consider the curved triangle PQR, made up of

part of the imaginary axis and parts of the circles. As k varies between 1 and $\sqrt{3}$, the angle at R will decrease continuously from $\frac{1}{3}\pi$ to 0. Thus there will be a sequence of points $Q_7, Q_8, Q_9, \ldots$, and corresponding points $R_7, R_8, R_9, \ldots$, such that the angles at R take the values $\frac{1}{7}\pi, \frac{1}{8}\pi, \frac{1}{9}\pi, \ldots$.

It is easy to see that the subgroup K_n of $G_1 \cup G$, generated by the operations of inversion in the circles PR_n, Q_nR_n and reflection in the line PQ_n is a discrete subgroup of $G \cup G_1$. Let $K_n \cap G = H_n$. Regarded as a group of transformations of the complex plane, it has as a fundamental domain the interior of the curved triangle PQ_nR_n, the reflection of this triangle in the line PQ_n, together with some of the boundary points of this region. It is one of the triangle groups well known in the theory of automorphic functions **(2; 3)**.

A H_n-fundamental domain in G is the set of all mappings t of G such that tP lies in the fundamental domain in the z-plane just described. For each n, the closure of the triangle PQ_nR_n lies in the interior of the upper half-plane, so G/H_n is compact.

The limit H of the sequence H_n has a fundamental domain which is obtained in the same way from the triangle $PQ_\infty R_\infty$, where $Q_\infty = i\sqrt{3}$, $R_\infty = -1$, and the R-angle of the curved triangle is zero. However, G/H is not compact because the closure of its fundamental domain contains the point R_∞, which is a boundary point of the upper half plane, and is not equal to tP for any $t \in G$.

Example 3. This example indicates that the conclusions of Theorem 2 cease to be true if G/H is not compact, even when G is connected and H finitely generated. The example was suggested to us in conversation by Professor Martin Kneser, and we are grateful to him for permission to include it here.

Let P, Q, R, S be four points on the real axis in the order indicated. Consider the operations t_1, t_2, t_3, t_4 of inversion in the circles on diameters SP, PQ, QR, RS. These generate a discrete subgroup H of $G \cup G_1$ which is a free product of four cyclic groups of order 2. Its fundamental domain in the upper half plane is the interior of the curved quadrilateral $PQRS$. Keep P, Q, S fixed and let R pass through a sequence of points tending to S. The group H will tend to a limit H_∞ which is generated by inversions in the circles SP, PQ, QS. The fundamental domain in the half-plane is the triangle PQS.

Now in the hyperbolic plane, the area of triangles with zero angles is a constant. Since the quadrilateral $PQRS$ is a union of two such triangles, its area is twice the area of the triangle PQS. Returning to the original group-space, we deduce without difficulty that

$$\mu(G/H \cap G) = 2\mu(G/H_\infty \cap G).$$

References

1. C. Chabauty, *Limite d'ensembles et geometrie des nombres*, Bull. Soc. Math. France, *78* (1950), 143–151.
2. L. R. Ford, *Automorphic functions* (New York: Chelsea, 1951).
3. R. Fricke and F. Klein, *Vorlesungen ueber die Theorie der Automorphen Funktionen* (Leipzig: Teubner, 1897–1912).
4. P. R. Halmos, *Measure theory* (New York: Van Nostrand, 1950).
5. A. M. Macbeath, *Abstract theory of packings and coverings, I* (to appear in Proc. Glasgow Math. Assoc.).
6. A. M. Macbeath and S. Świerczkowski, *On the set of generators of a subgroup*, Indag. Math., *21* (1959), 280–281.
7. K. Mahler, *On lattice points in n-dimensional star bodies. I, Existence Theorems*, Proc. Roy. Soc. London, Ser. *A.187* (1946), 151–187.
8. D. Montgomery and L. Zippin, *Topological transformation groups* (New York: Interscience tracts, 1955).
9. C. L. Siegel, *Discontinuous groups*, Ann. Math., *44* (1943), 674–678.
10. S. Świerczkowski, *Abstract theory of packings and coverings, II* (to appear in Proc. Glasgow Math. Assoc.).
11. A. Weil, *L'integration dans les groupes topologiques et ses applications* (Paris: Hermann, 1951).

Queen's College
Dundee, Scotland

CANONICAL FORMS FOR CERTAIN MATRICES UNDER UNITARY CONGRUENCE

J. W. STANDER AND N. A. WIEGMANN

1. Introduction. If A is a matrix with complex elements and if $A = A^T$ (where A^T denotes the transpose of A), there exists a non-singular matrix P such that $PAP^T = D$ is a diagonal matrix (see **(3)**, for example). It is also true (see the principal result of **(5)**) that for such an A there exists a unitary matrix U such that $UAU^T = D$ is a real diagonal matrix with non-negative elements which is a canonical form for A relative to the given U,U^T transformation. If $A = -A^T$, it is known (see **(3)** or **(4)**) that there exists a non-singular matrix P such that PAP^T is a direct sum of a zero matrix (if present) and of 2×2 blocks of the form:

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

The present work is concerned with the following. First, a canonical form is obtained for a complex skew-symmetric matrix under a U,U^T transformation where U is a unitary complex matrix; this form is analogous to that of the symmetric matrix mentioned above. Thereafter, matrices with real quaternion elements are considered. For such an A the *-transpose (denoted by A^*) is defined and is seen to be a generalization of the transpose (of a complex matrix) for the non-commutative case which at the same time retains the properties of the ordinary transpose in the commutative case. Quaternion matrices of the form $A = A^*$ and $A = -A^*$ are considered, in turn, and results analogous to those mentioned above for complex matrices are obtained which justify this generalization.

2. A normal form for a complex symmetric matrix under unitary congruence. To obtain this form the following is employed:

LEMMA 1. *If A is a complex, unitary, skew-symmetric matrix there exists a complex unitary matrix U such that $UAU^T = E$ is a direct sum of 2×2 matrices of the form*

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

It is evident that A must be of even order since it is skew-symmetric and non-singular. Let $A = A_1 + iA_2$, where A_1 and A_2 are real matrices, so that $A_1 = -A_1^T$ and $A_2 = -A_2^T$. Since $AA^{CT} = (A_1+iA_2)(A_1^T - iA_2^T) = I$,

Received April 23, 1959.

it follows that $A_1A_1^T+A_2A_2^T = I$ and $A_2A_1^T = A_1A_2^T$. The latter becomes $A_2A_1 = A_1A_2$. By a known theorem (see **(2)**, for example), there exists a real orthogonal matrix T such that $TA_1T^T = E_1$ and $TA_2T^T = E_2$ are direct sums of zeros and 2×2 matrices of the form

(i) $$\begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}$$

where $a > 0$ is real. Furthermore, it can be shown that, as in the present case, when A_1 and A_2 are both skew-symmetric, E_1 and E_2 can be regarded as conformable direct sums of 2×2 matrices of the above form, of 2×2 zero matrices, and of 1×1 zero matrices in such a way that whenever a single zero element appears in the direct sum of one, it appears in the same diagonal position in the other. (A 2×2 matrix of form (i) in one can correspond to a 2×2 zero matrix in the other, of course.) This may be seen as follows:

The statement is true or there is a first block (in E_1 or E_2) in the direct sum where it is not true; this would mean that there would be corresponding 3×3 diagonal blocks in E_1 and E_2, respectively, of the form

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & b \\ 0 & -b & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & a & 0 \\ -a & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where $a \neq 0$ and $b \neq 0$. But since $A_2A_1 = A_1A_2$, the above matrices must commute and they do not. Hence E_1 and E_2 can be considered to be direct sums which are conformable as described above.

Therefore $T(A_1+iA_2)T^T = E_1 + iE_2$ which is unitary (and non-singular); consequently, no 1×1 zero element can appear alone along the diagonal of E_1 and E_2 in the form described for each in the preceding paragraph. Therefore, E_1 and E_2 are each direct sums of 2×2 matrices of form (i) where $a \geqq 0$, so that E_1+iE_2 is a direct sum of 2×2 blocks of the form

(ii) $$E_0 = \begin{bmatrix} 0 & \alpha \\ -\alpha & 0 \end{bmatrix}$$

where α is non-zero complex. Since E_1+iE_2 is unitary, $\alpha\bar{\alpha} = 1$. Let $\alpha = e^{i\theta}$ and form the 2×2 unitary matrix

$$V = \begin{bmatrix} 0 & e^{-i\theta/2} \\ -e^{-i\theta/2} & 0 \end{bmatrix}.$$

Then VE_0V^T is a matrix of the form

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

If S is an appropriate direct sum of such V (determined from each 2×2 matrix in the direct sum E_1+iE_2), then $ST(A_1+iA_2)T^TS^T = E$, the direct

sum as described in the statement of the lemma, where $U = ST$ is a complex unitary matrix.

THEOREM 1. *If A is a complex skew-symmetric matrix, there exists a complex unitary matrix V such that $VAV^T = E \dot{+} 0$ where E is a direct sum of 2×2 matrices of the form*

$$\begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix},$$

where $a > 0$ is real; and conversely.

Let $A = HU = UK \neq 0$ be a polar representation of A where H and K are hermitian and U is unitary. (It may be noted that each $a > 0$ described in the statement of the theorem is actually a characteristic root of H or K). Since $A = HU = UK = -A^T = -U^TH^T = -K^TU^T$, and since the hermitian polar matrix H is unique, it follows from $A = HU = -K^TU^T$ that $H = K^T$ or $H = -K^T$ (since $-K^TU^T$ is also a polar form of A). But since K is positive definite, K^T is also, and $H = -K^T$ cannot hold (since H would not be positive definite). Therefore $H = K^T$.

If A, skew-symmetric, is non-singular, it must be of even order; in any event, the rank of A is even. If $A = HU$, the rank of $A =$ the rank of $H = r$, an even number.

For $H = K^T$ let V_1 be a complex unitary matrix such that $V_1HV_1^{CT} = D = D_0 \dot{+} 0$ (where 0 is absent if B is non-singular) where $D_0 = D_1 \dot{+} D_2 \dot{+} \ldots \dot{+} D_k$, where $D_i = d_iI_i$ is a real diagonal scalar matrix, $d_i \neq d_j$ for $i \neq j$, and $d_1 > d_2 > \ldots > d_k > 0$. If A is non-singular, it is known (see (9)) that the polar representation is unique, so that $A = HU = K^T(-U^T)$ implies that $U = -U^T$. If A is singular, this need not be true (8); as a matter of fact, it cannot be true if A is of odd order since U is non-singular.

Consider the case where $A = HU$ is singular. Let $V_1UV_1^{CT} = W$ and $V_1(-U^T)V_1^{CT} = W_1$; also let $V_1KV_1^{CT} = V_1H^TV_1^{CT} = M$. Then from

$$\begin{aligned} V_1AV_1^{CT} = V_1HUV_1^{CT} = V_1UKV_1^{CT} = V_1(-U^TH^T)V_1^{CT} \\ = V_1(-K^TU^T)V_1^{CT} \end{aligned}$$

it follows that $V_1AV_1^{CT} = DW = WM = W_1M = DW_1$. From $WM = W_1M$ it follows, in turn, that

$$W(V_1H^TV_1^{CT}) = W_1(V_1H^TV_1^{CT}),$$

or

$$WV_1V_1^TDV_1^CV_1^{CT} = W_1V_1V_1^TDV_1^CV_1^{CT},$$

so that $WV_1V_1^TD = W_1V_1V_1^TD$. Since $DW = DW_1$ (and since D has rank r), W and W_1 have like first r rows, and so $WV_1V_1^T$ and $W_1V_1V_1^T$ also have like first r rows; and from the last result in the preceding, $WV_1V_1^T$ and $W_1V_1V_1^T$ also have like first r columns. Let $WV_1V_1^T$ be of the form

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & X \end{bmatrix}$$

where A_{11} is an $r \times r$ matrix. Since $DW = W_1M = W_1V_1V_1^{\mathrm{T}}DV_1^{\mathrm{C}}V_1^{\mathrm{CT}}$, therefore $DWV_1V_1^{\mathrm{T}} = W_1V_1V_1^{\mathrm{T}}D$. From this relation it follows, after equating corresponding elements and noting that $W_1V_1V_1^{\mathrm{T}}$ is of the same form as $WV_1V_1^{\mathrm{T}}$ except for X, that A_{12} and A_{21} are zero matrices. Then:

$$WV_1V_1^{\mathrm{T}} = A_{11} \dot{+} X, \qquad W_1V_1V_1^{\mathrm{T}} = A_{11} \dot{+} Y,$$

$$W = (A_{11} \dot{+} X)V_1^{\mathrm{C}}V_1^{\mathrm{T}} = V_1UV_1^{\mathrm{CT}}, W_1 = (A_{11} \dot{+} Y)V_1^{\mathrm{C}}V_1^{\mathrm{CT}} = V_1(-U^{\mathrm{T}})V_1^{\mathrm{CT}},$$

$$U = V_1^{\mathrm{CT}}(A_{11} \dot{+} X)V_1^{\mathrm{C}}, -U^{\mathrm{T}} = V_1^{\mathrm{CT}}(A_{11} \dot{+} Y)V_1^{\mathrm{C}}.$$

Therefore, $U^{\mathrm{T}} = V_1^{\mathrm{CT}}(A_{11}^{\mathrm{T}} \dot{+} X^{\mathrm{T}})V_1^{\mathrm{C}} = V_1^{\mathrm{CT}}(-A_{11} \dot{+} [-Y])V_1^{\mathrm{C}}$ and so $A_{11} = -A_{11}^{\mathrm{T}}$ and A_{11} must also be unitary (since U^{T} is) and $Y = -X^{\mathrm{T}}$ where X is unitary but otherwise arbitrary. So $V_1UV_1^{\mathrm{T}} = A_{11} \dot{+} X$ and $V_1(-U^{\mathrm{T}})V_1^{\mathrm{T}} = A_{11} \dot{+} Y$.

Then $V_1AV_1^{\mathrm{T}} = V_1HV_1^{\mathrm{CT}}V_1UV_1^{\mathrm{T}} = V_1(-U^{\mathrm{T}})V_1^{\mathrm{T}}V_1^{\mathrm{C}}H^{\mathrm{T}}V_1^{\mathrm{T}} = (D_0 \dot{+} 0) \cdot (A_{11} \dot{+} X) = (A_{11} \dot{+} Y)(D_0 + 0)$. This means that $V_1AV_1^{\mathrm{T}} = D_0A_{11} \dot{+} 0$ where $D_0A_{11} = A_{11}D_0$ is of (even) order r, and A_{11} is unitary and skew-symmetric. It follows that $A_{11} = A_1 \dot{+} A_2 \dot{+} \ldots \dot{+} A_k$, where A_i is of the order of D_i in $D_0 = D_1 \dot{+} D_2 \dot{+} \ldots \dot{+} D_k$, and that each A_i is unitary and skew-symmetric and hence of even order. From the lemma for each A_i there exists a complex unitary U_i such that $U_iA_iU_i^{\mathrm{T}}$ is a direct sum of the 2×2 matrices described in the lemma. If $U = U_1 \dot{+} \ldots \dot{+} U_k$, then $UV_1AV_1^{\mathrm{T}}U^{\mathrm{T}} = D_0E_0 \dot{+} 0$ where E_0 is a direct sum of 2×2 matrices of the form described in the lemma. Then D_0E_0 is the matrix E described in the theorem, and since UV_1 is unitary, the theorem has been obtained. If A is non-singular, the same proof holds and $D = D_0$, $U = V_1^{\mathrm{CT}}A_{11}V_1^{\mathrm{C}}$, etc., and 0 does not appear in the final form $E \dot{+} 0$.

The converse is immediate.

3. A normal form for a *-symmetric quaternion matrix under unitary congruence. If two matrices A and B have elements which lie in a non-commutative domain, among the properties of the transpose which do not hold (as they do in the commutative case) is that $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$. If a matrix A with real quaternion elements is written in the form $A = A_1 + jA_2$ (where A_1 and A_2 are complex matrices), then $A^{\mathrm{T}} = A_1^{\mathrm{T}} + jA_2^{\mathrm{T}}$. Also, by the conjugate transpose of A is meant the matrix $A^{\mathrm{CT}} = A_1^{\mathrm{CT}} + (jA_2)^{\mathrm{CT}} = A_1^{\mathrm{CT}} - jA_2^{\mathrm{T}}$ (where A_1^{CT} denotes the complex conjugate transpose of A).

If the *-transpose of the matrix A is defined to be the matrix $A^* = A_1^{\mathrm{T}} + A_2^{\mathrm{T}}j$, it is seen that this includes the ordinary transpose of a complex matrix as a special case. Among the properties of the *-transpose which can easily be verified are the following: $(A^*)^* = A$; $A^* = ij\,A^{\mathrm{CT}}ji$; $(A + B)^* = A^* + B^*$; $(AB)^* = B^*A^*$; if A is non-singular, $(A^*)^{-1} = (A^{-1})^*$; $(A^*)^{\mathrm{CT}} = (A^{\mathrm{CT}})^*$. Define A to be *-symmetric if $A = A^*$, and to be *-skew-symmetric if $A = -A^*$. In the following, canonical forms are found for such matrices

under unitary congruence which are clearly generalizations of the theorems for the complex case stated in the two preceding sections.

The following lemma is first obtained:

LEMMA 2. *If U is a unitary quaternion matrix (that is, $UU^{CT} = I = U^{CT}U$) which is also *-symmetric ($U = U^*$), there exists a complex unitary matrix Z such that $ZUZ^T = D_0 + jD$ where D_0 and D are real diagonal matrices for which $D_0^2 + D^2 = I$.*

Let $U = U_1 + jU_2$, where U_1 and U_2 are complex matrices. Since $U = U_1 + jU_2 = U^* = U_1^T + U_2^T j$, it follows that $U_1 = U_1^T$ and $U_2 = U_2^{CT}$. Since, also, $UU^{CT} = (U_1 + jU_2)(U_1^{CT} - jU_2^T) = I$, $U_1U_1^{CT} + U_2^C U_2^T = I$ and $U_2U_1^{CT} = U_1^C U_2^T$ or, taking conjugates, $U_2^C U_1^T = U_1U_2^{CT}$ or $U_2^C U_1 = U_1U_2$. Let V be a complex unitary matrix such that $VU_2V^{CT} = D = D_1 \dotplus D_2 \dotplus \ldots \dotplus D_k$, where $D_t = d_tI_t$ for d_t real, $d_t \neq d_j$ for $i \neq j$, and where $d_1 > d_2 > \ldots > d_k$; also let $V^C U_1 V^{CT} = N$. Since $U_2^C U_1 = U_1U_2$, $V^C U_2^C V^T V^C U_1 V^{CT} = V^C U_1 V^{CT} VU_2V^{CT}$ or $DN = ND$. Therefore $N = N_1 \dotplus N_2 \dotplus \ldots \dotplus N_k$ is a direct sum conformable to D. Since $N = N^T$, $N_t = N_t^T$ for all i; consequently, there is a complex unitary W_t for each N_t such that $W_tN_tW_t^T = D_{1t}$ is a real diagonal matrix. If $W = W_1 \dotplus W_2 \dotplus \ldots \dotplus W_k$, then $WNW^T = D_{11} \dotplus D_{12} \dotplus \ldots \dotplus D_{1k} = D_0$ is a direct sum of real diagonal matrices. Then $WV^C(U_1 + jU_2)V^{CT}W^T = W(N + jD)W^T = D_0 + jD$ where D_0 and D are real diagonal matrices and WV^C is a complex unitary matrix. Furthermore, since U, V, and W are each unitary, $D_0 + jD$ is also and $(D_0 + jD)(D_0 - jD) = D_0^2 + D^2 = I$; the lemma is then true (and the converse is also, incidentally).

THEOREM 2. *If A is a *-symmetric quaternion matrix, there exists a quaternion unitary matrix U such that $UAU^* = D$ is a real diagonal matrix with non-negative diagonal elements; and conversely.*

This is clearly an analogue of the theorem for the complex case mentioned in §1, above; and its proof proceeds as does the proof for the complex case given in **(7**, p. 36**)**. If $A = HV = VK$ is the polar form of the quaternion matrix A (see **(6)**), the proof follows the same pattern except that *-transpose replaces T-transpose and the elements involved are quaternion (though the matrix D is still a real diagonal matrix). It is then found that for $A = HV = VH^*$, there exists **(7**, p. 37**)** a quaternion unitary matrix U such that $UAU^* = UHU^{CT}UVU^* = UVU^*(U^*)^{CT}H^*U^* = DW = WD$ where D is a real diagonal matrix as there described and $W = UVU^* = W^*$ is now a quaternion unitary matrix. Since D is real diagonal with like roots arranged together along the diagonal, $W = W_1 \dotplus W_2 \dotplus \ldots \dotplus W_t$ is a direct sum conformable to that of D (as a direct sum of scalar matrices) and each $W_t = W_t^*$ is unitary; it may be noted that if $D = D_1 \dotplus 0$ (as in **(7)**) and if 0 is present, W_t will be chosen to have these properties also. By the preceding lemma, a complex unitary Z_t can be chosen so that $Z_tW_tZ_t^* = Z_tW_tZ_t^T = D_{0t} + jD_{1t}$ where

D_{0i} and D_{1i} are real diagonal with the properties given. If $Z = Z_1 \dotplus Z_2 \dotplus \ldots \dotplus Z_t$, then $ZUAU^*Z^* = ZDWZ^* = DZWZ^* = D(D_a + j\,D_b) = D_c + j\,D_d$ where D_c and D_d are real diagonal and ZU is a quaternion unitary matrix.

To obtain the form given in the theorem, an additional step is required. If $\alpha = a + i\,b$, a and b real, is any complex number, since it is a 1×1 matrix and is equal to its transpose, there exists a complex unitary (number) $u = u_1 + i\,u_2$ so that $u\alpha u^{\mathrm{T}} = u\alpha u = r$, a real number. If j replaces i in this relation, the result still holds (since only j and real numbers are involved); therefore, if $\alpha = a + j\,b$ is any diagonal element of $D_c + j\,D_d$, there exists a quaternion unitary $u = u_1 + j\,u_2$ so that $u\alpha u^* = r$ is real. If this is applied to each diagonal element, the form described in the theorem can be obtained under the transformation required.

The converse follows immediately and the form is a canonical form, the diagonal elements being the characteristic roots of the hermitian polar matrix of A.

4. A normal form for a *-skew-symmetric matrix under unitary congruence. For this case there is the following lemma:

LEMMA 3. *If A is a *-skew-symmetric, unitary quaternion matrix, there exists a unitary complex matrix V such that VAV^{T} is a direct sum of 1×1 matrices of the form $+j\,i$ and $-j\,i$, and of 2×2 matrices of the form*

$$\begin{bmatrix} jri & a \\ -a & -jri \end{bmatrix}$$

where $a^2 + r^2 = 1$ and $a > 0$ and r are real numbers.

Since $A = A_1 + jA_2 = -A^* = -(A_1^{\mathrm{T}} + A_2^{\mathrm{T}} j)$, it follows that $A_1 = -A_1^{\mathrm{T}}$ and $A_2 = -A_2^{\mathrm{CT}}$. Since $AA^{\mathrm{CT}} = I = A^{\mathrm{CT}}A$, it follows, among other relations, that $A_2A_1^{\mathrm{CT}} = A_1^{\mathrm{C}}A_2^{\mathrm{T}}$ and $A_1^{\mathrm{T}}A_2 = A_2^{\mathrm{T}}A_1$. Since A_2 is skew-hermitian, let U be a complex unitary matrix such that $UA_2U^{\mathrm{CT}} = D = D_1 \dotplus D_2 \dotplus D_3 \dotplus \ldots \dotplus D_k$ is a direct sum of $D_s = ir_sI_s$ (where r_s is real), that is, of pure imaginary scalar matrices, arranged as follows: $r_s \neq r_t$ if $s \neq t$; if ir_s and $-ir_s$ are roots of A_2, their corresponding blocks appear successively on the diagonal; all such successive pairs of blocks, if present, appear first in D; and $D_k = 0$ if 0 is a root of A_2. Let $U^{\mathrm{C}}A_1U^{\mathrm{CT}} = M$.

From $A_2A_1^{\mathrm{CT}} = A_1^{\mathrm{C}}A_2^{\mathrm{T}}$ it follows that

$$UA_2U^{\mathrm{CT}}UA_1^{\mathrm{CT}}U^{\mathrm{T}} = UA_1^{\mathrm{C}}U^{\mathrm{T}}U^{\mathrm{C}}A_2^{\mathrm{T}}U^{\mathrm{T}},$$

or $DM^{\mathrm{CT}} = M^{\mathrm{C}}D^{\mathrm{T}}$; taking conjugates, $D^{\mathrm{C}}M^{\mathrm{T}} = MD^{\mathrm{CT}}$ or $-D(-M) = M(-D)$ or $DM = -MD$ (since $M^{\mathrm{T}} = -M$). Therefore, $D^2M = DDM = -D(MD) = MD^2$. Let $D = (D_1 \dotplus D_2) \dotplus \ldots \dotplus (D_{t-1} \dotplus D_t) \dotplus D_{t+1} \dotplus \ldots \dotplus D_k$ where the parentheses contain the successive pairs described earlier. Then $M = M_{12} \dotplus \ldots \dotplus M_{t-1,t} \dotplus M_{t+1} \dotplus \ldots \dotplus M_k$ where M_{rs} is of the

dimension of $D_r \dotplus D_s$, M_t is of the dimension of D_t, and all M_{rs} and M_t are complex skew-symmetric (since M is). Furthermore, since $-DM = MD$, it follows that $-(D_r \dotplus D_s)M_{rs} = M_{rs}(D_r \dotplus D_s)$ and $-D_tM_t = M_tD_t$ for all M_{rs} and M_t involved. Finally, it may be noted that $U^{C}AU^{CT} = U^{C}(A_1 + jA_2)U^{CT} = U^{C}A_1U^{CT} + jUA_2U^{CT} = M + jD$ must be *-skew symmetric and unitary. (Note that U is complex and $U^{C}A(U^{C})^{*} = U^{C}AU^{CT}$ is *-skew symmetric since A is also.)

(a) Consider, first, any relation $-(D_r \dotplus D_s)M_{rs} = M_{rs}(D_r \dotplus D_s)$ and, for convenience, the case where $r = 1$ and $s = 2$. Let $D_1 \dotplus D_2 = riI_1 \dotplus (-ri)I_2$ where I_1 and I_2 are, respectively, $p \times p$ and $q \times q$ identity matrices, $r \neq 0$, and assume, for specificity, that $p \leqq q$. Let M_{12} be of the form

$$\begin{bmatrix} M_1 & M_3 \\ -M_3^{T} & M_2 \end{bmatrix}$$

where M_1 and M_2 are, respectively, $p \times p$ and $q \times q$ matrices. From the relation $-(D_1 \dotplus D_2)M_{12} = M_{12}(D_1 \dotplus D_2)$, it follows that M_2 and M_1 are zero matrices (since $r \neq 0$). Now M_3 may be a zero matrix or it may not; before proceeding further, consider the latter case.

If M_3, a $p \times q$ matrix, is not zero, by a theorem of Eckert and Young **(1)** it follows that there exist complex unitary matrices V and W, of orders $p \times p$ and $q \times q$, respectively, such that $VM_3W = D$ is a $p \times q$ diagonal matrix with non-negative real elements (at least one of which is not 0 here) along the diagonal. (A $p \times q$ matrix is diagonal if the only non-zero elements are of the form a_{ii}.) Form the matrix

$$X = \begin{bmatrix} 0 & W^{T} \\ V & 0 \end{bmatrix}$$

which is complex unitary. Then $X(M_{12} + jD_{12})X^{T} = XM_{12}X^{T} + jX^{C}D_{12}X^{T}$ is a matrix of the form

$$\begin{bmatrix} 0 & -D^{T} \\ D & 0 \end{bmatrix} + j\begin{bmatrix} D_2 & 0 \\ 0 & D_1 \end{bmatrix}$$

where D is the above-mentioned $p \times q$ diagonal matrix. Let $N_1 = XM_{12}X^{T}$ and $N_2 = X^{C}D_{12}X^{T}$, and note that the dimension of $D_2 = q \geqq p$ = dimension of D_1, that D_1 and D_2 have non-0 diagonal elements, and D has at least one non-zero diagonal element; also, let the non-0 diagonal elements of D appear first along the diagonal. Consider N_1 and N_2 and perform the following operations on them: interchange the $q + 1$st column of N successively with the qth, $q - 1$st, $q - 2$nd, . . . , 2nd so that the $q + 1$st column becomes the second column and all succeeding columns are in the same order as before; and also perform the same row operations. This can be accomplished by a real orthogonal simularity transformation and there result from N_1 and N_2, respectively, the matrices

$$\begin{bmatrix} 0 & a_1 & 0 & 0 \\ -a & 0 & 0 & 0 \\ 0 & 0 & 0 & -D_3^{\mathrm{T}} \\ 0 & 0 & D_3 & 0 \end{bmatrix} \qquad \begin{bmatrix} -ri & 0 & 0 & 0 \\ 0 & ri & 0 & 0 \\ 0 & 0 & -riI_3 & 0 \\ 0 & 0 & 0 & riI_4 \end{bmatrix}$$

where I_3 and I_4 are, respectively, identity matrices of order $q-1$ and $p-1$, respectively. If the same procedure is applied to the lower right blocks (ignoring the first two rows and columns of each), it can be seen that a series of such steps provides a real orthogonal matrix Y such that the matrix $YX(M_{12}+jD_{12})X^{\mathrm{T}}Y^{\mathrm{T}}$ is a direct sum of 2×2 blocks of the form

$$\begin{bmatrix} -jri & a_i \\ -a_i & jri \end{bmatrix}$$

(where a_i and r are non-zero and real), and of single elements $-jri$ and $+jri$. But since YX is complex unitary, so is this direct sum, and so each 2×2 block and jri must be unitary. This means that $r^2+a_i^2=1$ and $r^2=1$; but since $a_i\neq 0$, this can only mean that jri and $-jri$ cannot appear singly in the direct sum. Therefore $YX(M_{12}+jD_{12})X^{\mathrm{T}}Y^{\mathrm{T}}$ is a direct sum of 2×2 blocks of the above form where $r^2+a_i^2=1$, $r\neq 0$ and $a_i\neq 0$. (If in the above $p\geqq q$, the roles of $+jri$ and $-jri$ are interchanged, but a simple (and allowable) operation at the close can still place the element $-jri$ in the $1-1$ position.)

All of the above in (a) occurs if M_3 is not a zero matrix. If $M_3=0$, then $M_{12}+jD_{12}=jD_{12}=j(D_1\dot{+}D_2)$ which is a direct sum with diagonal elements $+jri$, $r^2=1$; in this case no X and Y are required.

Therefore in $UAU^{\mathrm{CT}}=M+jD$, each $M_{rs}+j(D_r\dot{+}D_s)$ can be treated as above depending on whether or not M_{rs} is a zero matrix.

(b) Consider any relation $-D_iM_i=M_iD_i$ where D_i is a non-0 pure imaginary scalar matrix. Then $M_i=-M_i$ so M_i is a zero matrix and $M_i+jD_i=jD_i$ which has diagonal elements jri, $r^2=1$.

(c) If $D_k=0$ is present in $UA_2U^{\mathrm{CT}}=D$, then $M_k+jD_k=M_k=-M_k{}^{\mathrm{T}}$ a complex unitary matrix. By Lemma 1 there exists a complex unitary matrix U such that $UAU^{\mathrm{T}}=E$ is a direct sum as described in the lemma.

If the results of (a), (b), and (c) are combined, it is evident that a complex unitary matrix W can be constructed so that $WU^{\mathrm{C}}AU^{\mathrm{CT}}W^{\mathrm{T}}=W(M+jD)W^{\mathrm{T}}$ is a direct sum of 2×2 matrices of the form

$$\begin{bmatrix} jri & a \\ -a & -jri \end{bmatrix}$$

(where $a^2+r^2=1$, $a>0$ and r are real) and of 1×1 matrices of the form ji and $-ji$.

THEOREM 3. *If A is a *-skew-symmetric quaternion matrix, there exists a quaternion unitary matrix V such that $VAV^{*}=E\dot{+}0$ where E is a direct*

sum of 1×1 *matrices of the form* kji *and* $-kji$, $k > 0$ *real, and of* 2×2 *matrices of the form*

$$\begin{bmatrix} sji & t \\ -t & -sji \end{bmatrix}$$

where $t > 0$ *and* s *are real.*

The proof follows the pattern of that of Theorem 1. If $A = 0$, the result is trivial. If $A \neq 0$, let $A = HU = UK$ be a polar representation of A. If *-transpose replaces T-transpose in the earlier proof, it is evident that $H = K^*$. Here, however, the rank of a *-skew-symmetric matrix is not necessarily even (as the preceding lemma shows). If the earlier proof is followed, it is seen eventually that, using the same letters, $U = V_1^{CT}(A_{11} \dot{+} X)V_1^{*CT}$ and $U^* = -V_1^{CT}(A_{11} \dot{+} Y)V_1^{*CT}$ so that $U^* = V_1^{CT}(A_{11}^* \dot{+} X^*)V_1^{CT} = -V_1^{CT}(A_{11} \dot{+} Y)V_1^{*CT}$ and, since $V_1^{CT*} = V_1^{*CT}$, $A_{11}^* = -A_{11}$ is quaternion unitary. Then $V_1AV_1^* = V_1HV_1^{CT}V_1UV_1^* = (D_1 \dot{+} 0)(A_{11} \dot{+} X) = (D_1A_{11} \dot{+} 0) = V_1(-U^*)V_1^*V_1^{*CT}H^*V_1^* = (A_{11} \dot{+} Y)(D_1 \dot{+} 0) = (A_{11}D_1 \dot{+} 0)$. Since $D_1A_{11} = A_{11}D_1$, A_{11} is a direct sum, $A_1 \dot{+} A_2 \dot{+} \ldots \dot{+} A_k$, (of *-skew-symmetric, unitary quaternion matrices) conformable to the direct sum of D_1. For each A_i there exists, by the preceding lemma, a complex unitary matrix W_i so that $W_iA_iW_i^T$ has the form described in the lemma. If $W = W_1 + W_2 \dot{+} \ldots \dot{+} W_k \dot{+} I$ (where I is of the order of 0 in $D_1 \dot{+}0$), $WV_1AV_1^*W^T$ is then a direct sum of 1×1 matrices of the form kji and $-kji$ ($k > 0$ is real), of 2×2 matrices of the form

$$\begin{bmatrix} jrci & ac \\ -ac & -jrci \end{bmatrix}$$

where $ac > 0$ is real, and of a zero matrix. (WV_1 is a unitary quaternion matrix.)

References

1. C. Eckert and G. Young, *A principal axis transformation for non-hermitian matrices*, Bull. Amer. Math. Soc., *45* (1939), 118–121.
2. N. Jacobson, *Lectures in abstract algebra* (New York: D Van Nostrand, 1953), 184.
3. C. C. MacDuffee, *The theory of matrices* (Chelsea, 1946).
4. S. Perlis, *Theory of matrices* (Cambridge, 1952).
5. I. Schur, *Ein Satz ueber quadratische Formen mit komplexen Koeffizienten*, Amer. J. Math., *67* (1945), 472.
6. N. A. Wiegmann, *Some theorems on matrices with real quaternion elements*, Can. J. Math., *7* (1955), 191–201.
7. ——— *On unitary and symmetric matrices with real quaternion elements*, Can. J. Math., *8*, (1954), 32–39.
8. J. Williamson, *A polar representation of singular matrices*, Bull. Amer. Math. Soc., *41* (1935), 118–123.
9. A. Wintner and F. D. Murnaghan, *On a polar representation of non-singular square matrices*, Proc. Nat. Acad. Sci., U.S.A , *17* (1931), 676–678

Catholic University
Washington, D.C.

ON NILPOTENT PRODUCTS OF CYCLIC GROUPS

RUTH REBEKKA STRUIK

Introduction. In this paper $G = F/F_n$ is studied for F a free product of a finite number of cyclic groups, and F_n the normal subgroup generated by commutators of weight n. The case of $n = 4$ is completely treated (F/F_2 is well known; F/F_3 is completely treated in **(2)**); special cases of $n > 4$ are studied; a partial conjecture is offered in regard to the unsolved cases. For $n = 4$ a multiplication table and other properties are given.

The problem arose from Golovin's work on nilpotent products (**(1)**, **(2)**, **(3)**) which are of interest because they are generalizations of the free and direct product of groups: all nilpotent groups are factor groups of nilpotent products in the same sense that all groups are factor groups of free products, and all Abelian groups are factor groups of direct products. In particular (as is well known) every finite Abelian group is a direct product of cyclic groups. Hence it becomes of interest to investigate nilpotent products of finite cyclic groups.

Golovin has done this (as well as other things) in **(2)** and **(3)**. In **(2)** there are results for the first nilpotent product (metabelian product) and in **(3)** there is a unique decomposition theorem for nilpotent products of finite cyclic groups.

It might be conjectured that all finite nilpotent groups are nilpotent products of cyclic groups. However, in **(2)** and **(3)** Golovin notes examples of non-Abelian groups with $((G, G), G) = 1$ which are not of this form. Here it is shown that the Burnside group with exponent 3 (with three or more generators) is not of this form.

To be more precise, and using Golovin's notation: Let

$$F = \prod_{i=1}^{t}{}^{*} A_i$$

be the free product of the A_i. Let $(a, b) = a^{-1}b^{-1}ab$ and $(A, B) = \{(a, b) | a \in A, b \in B\}$ where A and B are subgroups of a group. Let $(A_i) = \{(A_i, A_j) | i \neq j\}$ where the A_i are considered as subgroups of F (the i in (A_i) is to indicate that it is formed from the A_i in F). Let $_0(A_i)_F$ be the normal subgroup generated by (A_i) in F, $_k(A_i)_F = (_{k-1}(A_i)_F, F)$. Then according to Golovin **(1)**, the kth nilpotent product of the A_i is

$$G = A_1(k)A_2(k) \ldots (k)A_t = F/_k(A_i)_F.$$

(If the A_i are cyclic, then $G = F/F_{k+2}$.)

From now on, Golovin's notation will be dropped.

Received October 17, 1958; in revised form January 21, 1960.

In **(6)** it is shown that if F is a free group with a finite number of generators, then every element of F/F_n can be uniquely expressed as a product of standard commutators. Here it is shown that if F is replaced by a free product of cyclic groups, then Hall's results hold "essentially" provided that all primes appearing in the orders of the factors are $\geq n-1$. If the primes are $< n-1$, then the situation is complicated. The case $n=4$ is completely treated here (that is, $p=2, n=4$); partial results and conjectures are offered for $n>4$ and $p<n-1$.

Section 1 gives preliminary results. In § 2, the "well-behaved" case ($p \geq n-1$) is handled, and in § 3, the other cases are discussed.

The author would like to thank W. Magnus for encouragement while preparing this paper, and R. Ree for reading the manuscript and for helpful criticisms. The author is also indebted to the referee for many improvements.

1. Preliminaries. Let G be an arbitrary group. As usual, $(a,b) = a^{-1}b^{-1}ab$ for $a, b \in G$ and if A, B are subgroups of G, then $(A,B) = \{(a,b) | a \in A, b \in B\}$. The lower central series of G is an infinite sequence of subgroups, $G_1, G_2, \ldots$, where $G_1 = G$, $G_2 = (G,G), \ldots, G_{n+1} = (G_n, G)$. $(((a_1,a_2),a_3), \ldots, a_n)$ will often be abbreviated $(a_1, \ldots, a_n)$. An element of the form $(((a_1,a_2),(a_3,a_4)), \ldots, a_n)$ (that is, with arbitrary arrangement of parentheses) will often be referred to as a commutator (of weight n), as opposed to a member of G_n which is (in general) a product of commutators (of weight n or greater). In this paper, F will stand for a free product of a finite number of cyclic groups: $F = \prod^* A_i$, A_i cyclic. (A_i may be finite or infinite). The following identities are often useful:

$$(1) \qquad \begin{aligned} (xy,z) &= (x,z)((x,z),y)(y,z) \\ (x,yz) &= (x,z)(z,(y,x))(x,y) \end{aligned}$$

In **(6)**, the following theorem is proved:

THEOREM H1. *Let F be a free group with t generators, $u_1, u_2, \ldots, u_t$. Let $u_1, \ldots, u_s$ be a sequence of standard commutators of weight $< n$ (See* **(7)**.*) of non-decreasing weight. Then every element, g, of $F/F_n = G$ (free nilpotent group) can be uniquely expressed as*

$$g = \prod_{i=1}^{s} u_i^{c_i}$$

where the c_i are rational integers. If

$$h = \prod u_i^{d_i} \in G,$$

then

$$gh = \prod u_i^{e_i},$$

where $e_i = f_i(c_j, d_k)$ are polynomials with integer coefficients in the c_j and the d_k (for example, $e_i = c_i + d_i$; $1 \leq i \leq t$). If s-tuples of the form $(c_1, \ldots, c_s)$,

c_i rational integers are taken with multiplication given by $(c_1, \ldots, c_s) \times (d_1, \ldots, d_s) = (f_1(c_j, d_k), \ldots, f_s(c_j, d_k))$, the set of these s-tuples forms a nilpotent group isomorphic to F/F_n.

Throughout this paper, Hall's collection process will be frequently used. Several of its important theorems will now be summarized:

THEOREM H2: *Let R, S be any two elements of a group; let $u_1, u_2, \ldots,$ be a fixed sequence of commutators in R and S of non-decreasing weight, that is, $u_1 = (R, S)$, $u_2 = ((R, S), R)$, $u_3 = ((R, S), S)$, etc. Then*

$$(RS)^n = R^n S^n u_1^{f_1(n)} u_2^{f_2(n)} \ldots u_i^{f_i(n)} \ldots \tag{2}$$

where

$$f_i(n) = a_1\binom{n}{1} + a_2\binom{n}{2} + \ldots + a_{w_i}\binom{n}{w_i} \tag{3}$$

a_i are rational integers and w_i is the weight of u_i as a commutator in R and S. (2) is an identity if the group is nilpotent; otherwise (2) can be considered as giving a series of "approximations" to $(RS)^n$ modulo successive members of the lower central series.

The proof of Theorem H1 also gives

THEOREM H3. *Let $R_1, R_2, \ldots, R_s$ be any s elements of a group. Let $u_1, u_2, \ldots,$ be a fixed sequence of commutators in the R_i of non-decreasing weight (weight > 2). Let $i_1, i_2, \ldots, i_s$ be any fixed permutation of $1, 2, \ldots, s$. Then*

$$(R_1 R_2 \ldots R_s)^n = R_{i_1}^n R_{i_2}^n \ldots R_{i_s}^n u_1^{f_1(n)} u_2^{f_2(n)} \ldots u_i^{f_i(n)} \ldots \tag{4}$$

where $f_i(n)$ are of form (3) with w_i the weight of u_i in the R_j.

From Theorem H1 we can obtain

LEMMA H1. *Let X, Y be any elements of a group, and let $u_1, u_2, \ldots,$ be any fixed sequence of commutators in X and (X, Y) of non-decreasing weight; then*

$$(X^n, Y) = (X, Y)^n u_1^{f_1(n)} u_2^{f_2(n)} \ldots u_i^{f_i(n)} \ldots \tag{5}$$

where the $f_i(n)$ are like (3) with w_i as the weight of u_i in X and (X, Y).

Proof of Lemma H1. (5) follows from (2) in view of

$$(X^n, Y) = X^{-n}Y^{-1}X^nY = X^{-n}[Y^{-1}XY]^n = X^{-n}[X(X, Y)]^n$$
$$= X^{-n}X^n(X, Y)^n u_1^{f_1(n)} \ldots = (X, Y)^n u_1^{f_1(n)} \ldots .$$

LEMMA H2. *Let α be a fixed integer and G a group such that $G_n = 1$. Then if $b_j \in G$ and $r < n$,*

$$(b_1, \ldots, b_{i-1}, b_i^\alpha, b_{i+1}, \ldots, b_r) = (b_1, \ldots, b_r)^\alpha v_1^{f_1(\alpha)} v_2^{f_2(\alpha)} \ldots \tag{6}$$

where the v_k are commutators in $b_1, \ldots, b_r$ of weight $> r$, and every b_j, $1 \leqslant j \leqslant r$ appears in each commutator v_k. The f_i are of form (3) where w_i is the weight of v_i minus $(r - 1)$.

Proof. (6) is (5) with $r = 2$, $i = 1$, and $\alpha = n$. For $r = 2$, $i = 2$, and $\alpha = n$, take inverses on each side of (5).

$$(Y, X^n) = u_s^{-f_s(\alpha)} \dots u_1^{-f_1(\alpha)} (Y, X)^\alpha \tag{7}$$

where $u_s \in G_{n-1}$. Since $G_n = 1$, s is finite. Now apply (4):

$$(Y, X^n) = u_s^{-f_s(\alpha)} \dots u_1^{-f_1(\alpha)} (Y, X)^\alpha \qquad [R_t = (Y, X) \text{ or } u_j^{-f_j(\alpha)}] \tag{8}$$
$$= [(Y, X)^\alpha u_1^{-f_1(\alpha)} \dots u_s^{-f_s(\alpha)}] w_1^{f_1(1)} w_2^{f_2(1)} \dots$$

where w_t are commutators in $(Y, X)^\alpha$ and $u_j^{-f_j(\alpha)}$. Use induction starting with $(Y, X) \in G_{n-1}$. For $(Y, X) \in G_{n-s}$, assume the theorem (that is, (6)) is true for commutators $\in G_{n-s+1}$, and use this and (1) to express w_j in desired form. One will obtain as exponents in the expansions, expressions of the form

$$\left(\binom{\alpha}{i} \right). \tag{9}$$
$$j$$

From its meaning in terms of the number of subsets of a set, (9) is an integral-valued function of α (of degree $i \times j$). By (3.21) p. 64 of **(5)**, this can be expressed in the form

$$a_1 \binom{\alpha}{1} + \dots + a_{i \times j} \binom{\alpha}{i \times j}$$

a_t rational integers. This is sufficient to show

$$(Y, X^n) = (Y, X)^\alpha \prod v_k^{f_k(\alpha)} \tag{10}$$

which completes the proof for $r = 2$.

Suppose true for r, then for

$$(c_1, c_2, \dots, c_{t-1}, c_t^\alpha, c_{t+1}, \dots, c_{r+1}) \qquad i > 2 \tag{11}$$

put $b_1 = (c_1, c_2)$, $b_t = c_{t+1}$, $i = 2, \dots, r$ in (6) and use induction hypothesis. For

$$(c_1^\alpha, c_2, \dots, c_{r+1}) \tag{12}$$

put

$$X = (c_1^\alpha, c_2, \dots, c_r), \; Y = c_{r+1}.$$

By induction

$$X = (c_1, \dots, c_r)^\alpha \prod w_k^{f_k(\alpha)}.$$

Now use (1) an appropriate number of times with

$$x, y, z = (c_1, \dots, c_r)^\alpha, w_k^{\binom{\alpha}{i}} \quad \text{or} \quad c_{r+1}$$

and the induction hypothesis to put

$$(X, Y) = ((c_1, \dots, c_r)^\alpha \prod w_k^{f_k(\alpha)}, c_{r+1}) \tag{13}$$

in the form of (6).

A similar proof holds for

$$(c_1, c_2^a, c_3, \ldots, c_{r+1}).$$

Throughout this proof, we have implicitly used the fact that an arbitrary commutator can be expressed as a product of commutators of the form $(b_1, \ldots, b_r)$. Or to express the same idea in a different way, (6) can be proved in the same way, if $(b_1, \ldots, b_{i-1}, b_i^a, b_{i+1}, \ldots, b_r)$ and $(b_1, \ldots, b_r)$ are replaced by arbitrary commutators (that is, monomial commutators with parentheses arranged arbitrarily).

Let gcd stand for greatest common divisor and $\gcd(\alpha_1, \ldots, \alpha_k)$ stand for the gcd of the rational integers $\alpha_1, \ldots, \alpha_k$. The $\gcd(\alpha_1, \ldots, \alpha_k, 0) = \gcd(\alpha_1, \ldots, \alpha_k)$. This should not be confused with $(a_1, \ldots, a_k)$, a member of G_k, G a group, since $a_i \in$ a group, and will not be rational integers (in this paper). A cyclic group of order 0 will be understood to be infinite cyclic.

LEMMA 1. *Let*

$$F = \prod_{i=1}^{t} * A_i,$$

A_i cyclic of order α_i. Let a_i generate A_i. Let $n \geqslant 3$ be a fixed positive integer, and let all primes appearing in the factorizations of the $\alpha_i \geqslant n-1$. Let $G = F/F_n$. If $v \in G$, and

$$v = (a_{i_1}, \ldots, a_{i_k}),$$

then $v^N = 1$, where

$$N = \gcd(\alpha_{i_1}, \ldots, \alpha_{i_k}), \quad k \geqslant 2$$

(some of the α_{i_j} (or a_{i_j}) may equal each other). If w is a product of commutators like v in which every commutator contains all the distinct a_i appearing in v, then $w^N = 1$. Hence $w^N = 1$ where w is an arbitrary commutator.

Proof. Let

$$v = (a_{i_1}, \ldots, a_{i_{n-1}}) \in G_{n-1}.$$

By (6)

$$(14) \quad 1 = (a_{i_1}, \ldots, a_{i_j}^{\alpha_{i_j}}, \ldots, a_{i_{n-1}}) = (a_{i_1}, \ldots, a_{i_{n-1}})^{\alpha_{i_j}} \prod v_m^{\alpha_{i_j}}$$

$$1 \leqslant j \leqslant n-1$$

where all $v_m = 1$ since $G_n = 1$. Hence the Lemma holds for $k = n-1$. Since G_{n-1} is Abelian, $w^N = 1$ if w is a product of commutators of weight $n-1$ in which the same a_i appear in each commutator.

Suppose true for $k+1$, that is, if

$$v = (a_{i_1}, \ldots, a_{i_{k+1}}),$$

then $v^N = 1$ where

$$N = \gcd(\alpha_{i_1}, \ldots, \alpha_{i_{k+1}}),$$

and if w is a product of commutators of weight $k+1$ or greater in

$$a_{i_1}, \ldots, a_{i_{k+1}},$$

then $w^N = 1$. Consider $(a_{i_1}, \ldots, a_{i_k})$. By (6)

$$(15)\quad 1 = (a_{i_1}, \ldots, a_{i_j}^{\alpha_{i_j}}, \ldots, a_{i_k}) = (a_{i_1}, \ldots, a_{i_k})^{\alpha_{i_j}} \prod v_m^{f_m(\alpha_{i_j})} \quad 1 < j < k.$$

$$\prod v_m^{f_k(\alpha_{i_j})} = 1$$

by the induction hypothesis, and the assumption on the primes; hence

$$(a_{i_1}, \ldots, a_{i_k})^{\alpha_{i_j}} = 1.$$

Hence

$$(a_{i_1}, \ldots, a_{i_k})^N = 1 \text{ where } N = \gcd(\alpha_{i_1}, \ldots, \alpha_{i_k}).$$

Making use of (4), one obtains that if w is the product of commutators of weight k or greater in $a_{i_1}, \ldots, a_{i_k}$, then $w^N = 1$. Note that every factor of w must contain all the distinct a_{i_j}, and that in a nilpotent group, every commutator can be expressed a product of commutators of the form $(a_{i_1}, \ldots, a_{i_k})$.

For the case $n = 4$, (5) becomes:

LEMMA 2. *If G is any group and $a, b \in G$, then*

$$(16)\quad (a^r, b^s) = (a,b)^{rs}((a,b),a)^{s\binom{r}{2}}((a,b),b)^{r\binom{s}{2}} \quad \mod G_4,$$

$$(b^r, a^s) = (a,b)^{-rs}((a,b),a)^{-r\binom{s}{2}}((a,b),b)^{-s\binom{s}{2}} \quad \mod G_4,$$

$$\text{where} \quad \binom{r}{2} = \frac{r(r-1)}{2}.$$

Lemma 2 is proved in **(14)** and is a particular case of (5) in which the $f_i(n)$ have been computed. The proof of (16) is based on the work of Magnus **(11)**.

2. The "well-behaved" case.

THEOREM 1. *Let A_1, A_2, A_3 be cyclic groups of orders $\alpha_1, \alpha_2, \alpha_3$ respectively, α_i odd integers. Let a_i generate A_i. Let*

$$F = \prod_{i=1}^{i} * A_i.$$

Let $u_1, \ldots, u_{14}$ be a sequence of standard monomial commutators of non-decreasing weight in a_1, a_2, a_3 of weight $\leqslant 3$. (See **(7)**.*) Let $N_i = \alpha_i$ if u_i is of weight 1; let $N_i = \gcd(\alpha_i, \alpha_j)$ if $u_i = (a_i, a_j)$, and let $N_i = \gcd(\alpha_i, \alpha_j, \alpha_k)$ if a_i, a_j, a_k appear in u_i of weight 3. Then every element of g of F/F_4 can be uniquely expressed as*

$$(17)\quad g = \prod u_i^{c_i}$$

where the c_i are integers modulo N_i. If

$$h = \prod u_i^{d_i}$$

is another element of F/F_4, then

$$gh = \prod u_i^{e_i}$$

where $e_i = f_i(c_j, d_k)$ are the polynomials with integral coefficients of Theorem H1.

(Theorem 1 is a generalization of a lemma appearing in **(15)**.)

Proof. By Lemma 1, $u_i^{N_i} = 1$. Hence every element of G can be expressed in the form of (17) where the c_i are integers modulo N_i. The problem is to show that this expression is unique.

Let $u_1, \ldots, u_{14}$ be $a_1, a_2, a_3, (a_1, a_2), (a_1, a_3), (a_2, a_3), (a_1, a_2, a_1), (a_1, a_3, a_1), (a_2, a_3, a_2), (a_1, a_2, a_2), (a_1, a_3, a_3), (a_2, a_3, a_3), (a_1, a_2, a_3), (a_2, a_3, a_1)$, respectively. If another sequence of standard commutators is chosen, a similar proof will hold. Since (a_i, a_j), $i \neq j$ generate (G, G) modulo G_3 and since

$$((a, b), c)((b, c), a)((c, a), b) = 1 \text{ modulo } G_4 \quad \text{(see (11))}$$

and (a_i, a_j, a_k) generate G_3 modulo G_4, the u_i specified above do form a basis for G. The following change of notation will be made:

let $u_{ij} = (a_i, a_j)$ and designate the corresponding c_i, d_i, e_i by c_{ij}, d_{ij}, e_{ij} respectively;

Let $u_{iji} = (a_i, a_j, a_i)$ and designate the corresponding c_i, d_i, e_i by c_{iji}, d_{iji}, e_{iji} where $i < j$;

let $u_{ijj} = (a_i, a_j, a_j)$ and designate the corresponding c_i, d_i, e_i by c_{ijj}, d_{ijj}, e_{ijj} where $i < j$;

let $u_{ijk} = (a_i, a_j, a_k)$ and designate the corresponding c_i, d_i, e_i by c_{ijk}, d_{ijk}, e_{ijk} where $i < j < k$;

let $u_{jki} = (a_j, a_k, a_i)$ and designate the corresponding c_i, d_i, e_i by c_{jki}, d_{jki}, e_{jki} where $i < j < k$.

For Theorem 1, u_{jki} and u_{ijk} are u_{231} and u_{123} respectively, but the more general notation is used here for the sake of Theorem 2.

Then a somewhat laborious computation gives

$$\begin{aligned}
e_i &= c_i + d_i \\
e_{ij} &= c_{ij} + d_{ij} - c_i d_i \\
e_{iji} &= c_{iji} + d_{iji} - c_j\binom{d_i}{2} + c_i d_i \\
e_{ijj} &= c_{ijj} + d_{ijj} - d_i\binom{c_j}{2} + c_i d_j - d_i d_j c_j \\
e_{ijk} &= c_{ijk} + d_{ijk} + c_{ik} d_j + c_i d_k - d_i c_j c_k - c_k d_i d_j - c_j d_i d_k \\
e_{jki} &= c_{jki} + d_{jki} + c_{ji} d_i + c_{ik} d_j - c_k d_i d_j.
\end{aligned} \tag{18}$$

Note that these are the $f_i(c_j, d_k)$ of Theorem H1 for $n = 4$, and the particular sequence of u_i chosen here. Also note that they apply unambiguously if they are interpreted as integers modulo the appropriate gcd. For example, e_{121} is an integer modulo $\gcd(\alpha_1, \alpha_2)$; c_2, d_1, and c_{12} appear in its formula, but since c_2, d_1, and c_{12} are integers modulo α_2, α_1, and $\gcd(\alpha_1, \alpha_2)$ respectively, no ambiguity arises in the computation of a particular e_{121}. By Theorem H1, if one takes 14-tuples, $(c_1, \ldots, c_{14})$, $(d_1, \ldots, d_{14})$, c_i, d_j rational integers and lets (18) define a multiplication, a group isomorphic to F/F_4 (free nilpotent group) (F a free group) is obtained. The same proof will go through if the c_i, d_j are integers modulo the appropriate gcd. (One can also check the group axioms directly, a tedious verification.) Note that α_i odd is essential here, since (18) involves

$$\binom{c_i}{2}, \quad \binom{d_i}{2},$$

and this will give difficulty if one is dealing with integers modulo an even integer.

THEOREM 2. *Let $A_1, \ldots, A_t$ be cyclic groups of order $\alpha_1, \ldots, \alpha_t$ respectively, α_i odd integers or 0. Let a_i generate A_i. Let*

$$F = \prod_{i=1}^{t} {}^{*} A_i.$$

Let $u_1, u_2, \ldots,$ be a sequence of standard (monomial) commutators of non-decreasing weight in the a_i of weight < 3 (see (7)). Let $N_i = \alpha_i$ if u_i is of weight 1; $N_i = \gcd(\alpha_i, \alpha_j)$ if $u_i = (a_i, a_j)$ and $N_i = \gcd(\alpha_i, \alpha_j, \alpha_k)$ if a_i, a_j, a_k appear in u_i (of weight 3). Then every element of F/F_4 can be uniquely expressed as

$$g = \prod u_i^{c_i}$$

where c_i are integers modulo N_i. (If $N_i = 0$, then c_i is a rational integer.) If

$$h = \prod u_i^{d_i}$$

is another element of F/F_4, then

$$gh = \prod u_i^{e_i}$$

where $e_i = f_i(c_j, d_k)$ are the polynomials with integral coefficients of Theorem H1.

Proof. The proof is the same as that of Theorem 1. (18) is a multiplication table for G provided the standard commutators are arranged in the order: a_i, (a_i, a_j), (a_i, a_j, a_i), (a_i, a_j, a_j), (a_i, a_j, a_k), (a_j, a_k, a_i) with $i < j < k$.

Comment. Since every finite nilpotent group is a direct product of prime power groups, the α_i may be assumed to be prime powers or 0.

COROLLARY 1. *Let*

$$g = \prod u_i^{c_i}$$

be a particular element of G. Then $g^N = 1$ where N is the least common multiple of the orders of the $u_i^{c_i}$ appearing in g unless $g \notin (G, G)$ and $3|N$. In the latter case, $g^{3N} = 1$, and g may be of order 3N. If any of the u_i appearing in g are infinite cyclic, then g is of infinite order.

The author is indebted to the referee for a simplification of the statement and proof of this corollary.

Proof. If $g \in (G, G)$, then since (G, G) is Abelian, the Corollary follows. If g contains a u_i which is infinite cyclic, then by (4) and the unique representation of g, g must be infinite cyclic. If $g \notin (G, G)$, and all the factors are of finite order, then at least one of the u_i is equal to an a_j. Looking at (4) with the n of (4) put equal to N, it is obvious that $g^N = 1$ (Lemma 1 is used here) provided $3 \nmid N$, since the $f_i(N)$ will involve N,

$$\binom{N}{2}, \text{ and } \binom{N}{3}.$$

(All commutators are of weight $\leqslant 3$).

If $3|N$, i.e., $3|\alpha_j$ for an a_j appearing in g, then the above reasoning indicates that $g^{3N} = 1$. g can actually be of order $3N$; for example, if

$$G = \{a, b \mid a^3 = b^3 = 1, \quad G_4 = 1\} \tag{19}$$

an actual computation shows that ab, ab^2, a^2b, and a^2b^2 are of order 9; in this case

$$(a^t b^t)^3 \in G_3. \tag{20}$$

Another way of seeing this is to consider equation (7) of **(14)** (due to Sanov) that is,

$$((a, b), b)^{\frac{1}{3}N} \in F(N)F_4 \tag{21}$$

where F can be any group generated by a and b and $F(N)$ is the normal subgroup generated by all $N = 3N'$ powers of elements of F. If a and b are of order N and if all elements of F/F_4 were of order N (or less), then $((a, b), b)$ would be of order $\leqslant \frac{1}{3}N$ and not N as Theorem 2 indicates (that is, $t = 2$, $\alpha_1 = \alpha_2 = N$).

Comment. The group G given by (19) is a kind of curiosity, for p-groups, since it is ***not*** regular in the sense of Hall **(5**, p. 73**)**. However all groups of the form

$$G = \{a, b | a^{p^n} = b^{p^n} = 1, G_4 = 1\} \tag{22}$$

with $p \geqslant 5$, p a prime, are regular groups in the sense of Hall.

A similar comment can be made in connection with

$$G = \{a, b \mid a^2 = b^2 = 1, G_3 = 1\}, \tag{23}$$

a group of order 8.

COROLLARY 2. *The group* $S_t = \{a_i | 1 \leqslant i \leqslant t, s^3 = 1, s \in S_t\}$ *is not a nilpotent product of cyclic groups of order three, except for* $t = 2$ *when* $S_2 = F/F_3$, $F = \{a_1\}^*\{a_2\}$. *However,* S_t *is a fully regular product* (see [1]) *of the* $\{a_i\}$, *and, in particular, it is the third Burnside product of the* $\{a_i\}$ **(12)**.

Proof. The only candidates for S_t to be a nilpotent product are the first (F/F_3) and second (F/F_4) nilpotent products. (F a free product of cyclic groups of order three.) Since $((a_1, a_2), a_1) \neq 1$ in F/F_4 while $((a_1, a_2), a_1) = 1$ in S_t (cf. **(9)**), S_t cannot be a second nilpotent product. As for the first nilpotent product (that is, F/F_3), $(a_1, a_2, a_3) = 1$ in this case, while $(a_1, a_2, a_2) \neq 1$ in S_t. However, if $t = 2$, $S_2 = F/F_3$ where F is the free product of two cyclic groups of order three, and $S_2 =$ first nilpotent product of $\{a_1\}$ and $\{a_2\}$ **(2, 9)**.

THEOREM 3. *Let* $A_1, \ldots, A_t$ *be cyclic groups of order* $\alpha_1, \ldots, \alpha_t$ *respectively. If* A_i *is infinite cyclic, let* $\alpha_i = 0$. *Let* a_i *generate* A_i; *let* $F = \prod^t_{i=1}{}^* A_i$. *Let* $n \geqslant 3$ *be a fixed positive integer and let all the primes appearing in the factorizations of the* $\alpha_i \geqslant n - 1$. *Let* $u_1, \ldots,$ *be a sequence of standard monomial commutators of non-decreasing weight in the* a_i *of weight* $\leqslant n - 1$. *Let* $N_i = \alpha_i$ *if* u_i *of weight* 1, *and*

$$N_i = \gcd(\alpha_{i_1}, \ldots, \alpha_{i_k}) \text{ if } a_{i_j},\ 1 \leqslant j \leqslant k,$$

appears in u_i. *Then every element* g, *of* $G = F/F_n$ *can be uniquely expressed as*

$$g = \prod u_i^{c_i}$$

where the c_i *are integers modulo* N_i. (*If* $N_i = 0$, c_i *is a rational integer.*) *If*

$$h = \prod u_i^{d_i}$$

is another element of F/F_n, *then*

$$gh = \prod u_i^{e_i}$$

where $e_i = f_i(c_j, d_k)$ *are the polynomials with integer coefficients of Theorem* H1.

We note that if F were free, the u_i of weight k would form a basis for F_k/F_{k+1}, see **(7)**.

Proof. The proof is exactly the same as that of Theorems 1 and 2. Lemma 1 shows that the orders of the u_i are as stated in the theorem, so that every element of g is of the form stated, and the only problem is uniqueness. As in Theorem 1, one can theoretically compute a multiplication table similar to (18). This is computed by multiplying

$$u_1^{c_1} \ldots u_s^{c_s} \cdot u_1^{d_1} \ldots u_s^{d_s} = u_1^{c_1} \ldots u_{s-1}^{c_{s-1}} u_1^{d_1} u_2^{c_2} (u_3^{c_3}, u_1^{d_1}) u_1^{d_2} \ldots$$

etc., and using (5), (6), or (10), or a suitable modification of them. The coefficients of the multiplication table will involve

$$c_i, d_j, \binom{c_i}{2}, \binom{d_j}{2}, \ldots, \binom{c_i}{n-2}, \binom{d_j}{n-2}.$$

Note that the $f_k(n)$ of largest order will come from applying (5) and (10) to

$$(u_s^{e_s}, u_1^{e_1}) \quad \text{or} \quad (u_j^{e_j}, u_i^{e_i}) \; i < j$$

and since in (5) one is dealing with commutators in X and (X, Y), the corresponding coefficients of the $f_i(c_j, d_k)$ will involve at most

$$\binom{c_i}{n-2}, \binom{d_j}{n-2} \quad \text{not} \quad \binom{c_i}{n-1}, \binom{d_j}{n-1}.$$

Hence, since all the primes of the $\alpha_j \geqslant n-1$, no ambiguity will occur because the c_i and d_j are taken modulo the appropriate gcds. Hence Theorem H1 can be used with the $f_i(c_j, d_k)$ considered as integers modulo the appropriate gcds, and this is sufficient to prove the theorem.

COROLLARY. *Let*

$$g = \prod u_i^{e_i}$$

be the unique representation of an element of G. Let N be the least common multiple of the orders of the $u_i^{e_i}$ appearing in g.

Case I. *If one of the u_i is infinite cyclic, then g is infinite cyclic.*

Case II. *All the primes appearing in the orders of the u_i are greater than $n-1$* OR *$g \in (G, G)$. (g is assumed to have factors which are all of finite order.) Then $g^N = 1$.*

Case III. *$g \notin (G, G)$ and p (a prime) $= n-1$ and p appears in the factorization of one of the α_j where a_j is a factor of g. Then $g^{pN} = 1$, and there are cases where $g^N \neq 1$.*

Proof. Case I follows from (4) and the uniqueness of the representation of g. (Consider what happens in (4) to the infinite cyclic u_i of least weight.) For Case II, consider (4) where $R_i = u_i^{e_i}$ (of g). If every u_i (of g) $\in G_{n-1}$, then (4) gives $g^N = 1$. If $g \in G_{n-s}$, use induction on s, (4), (6) Lemma 1, and the fact that the u_i of (4) can be expressed as products of commutators of the form

$$(a_{i_1}, \ldots, a_{i_k}).$$

If all the primes appearing in the α_j of u_i (of g) are greater than $n-1$, the $f_i(N)$ of (4) will involve

$$\binom{N}{2}, \ldots, \binom{N}{n-1},$$

and $N \mid f_i(N)$, hence

$$u_i^{f_i(N)} = 1$$

and $g^N = 1$. If $g \in (G, G)$, then the same proof holds except that the $f_i(N)$ involve

$$\binom{N}{2}, \ldots, \binom{N}{n-2}.$$

For Case III, if $p = n - 1$, p a prime, and for some a_j (appearing in g, $p|\alpha_j$ (hence $p|N$), then

$$\binom{N}{n-1} = \binom{N}{p}$$

may cause difficulty, but in any case,

$$pN \left| \binom{pN}{p} \right.$$

and hence $g^{pN} = 1$. If $\alpha_1 = \alpha_2 = \ldots = \alpha_t = p^\lambda = N$ where $p = n - 1$, then according to Sanov **(13)**,

$$\underbrace{(a_1, a_2, \ldots, a_2)}_{p-1 \text{ times}}^{p^{\lambda-1}} \in F(p^\lambda)F_{p+1} = F(p^\lambda)F_n \qquad (p = n - 1)$$

where

$$F(p^\lambda) = \{x^{p^\lambda} | x \in F\}.$$

If $g^{p^\lambda} = 1$ for every element of F/F_n,

$$\underbrace{(a_1, a_2, \ldots, a_2)}_{p-1 \text{ times}}$$

would have order $p^{\lambda-1}$ or less which contradicts Theorem 3 (according to which $(a_1, a_2, \ldots, a_2)$ has order p^λ). Hence there exist elements which have order $p^{\lambda+1} = pN$. In view of the Corollary to Theorem 2, probably

$$(a_1a_2)^{p^\lambda} \neq 1.$$

Comment. If $a_i^p = 1$, $1 \leqslant i \leqslant t$, $n \leqslant p$, p a prime, all elements of G are of order p, and hence G is a factor group of the Burnside group B with exponent p in t generators. In **(10)** and **(13)** it is shown that B_s/B_{s+1} has the same rank as F_s/F_{s+1} (F the free group with t generators) for $s = 1, 2, \ldots, p - 1$. This provides a partial verification of Theorem 3.

Comment. In **(4)** Gruenberg states and proves "Hall's Second Basis Theorem." It is essentially Theorem 3 for the case $\alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_t = p^\lambda$ and $n \leqslant p$. Theorem 3 shows that Hall's Second Basis Theorem holds "one step further" for $n = p + 1$.

3. The "ill-behaved" case. If $p < n - 1$, the proofs above break down The case of $A = \{a\}$, $B = \{b\}$, $a^2 = b^2 = 1$ is of interest. In $F = A*B$ (the free product of A and B), (A, B) is infinite cyclic and generated by (a, b). Since

(24) $$1 = (a, b^2) = (a, b)^2((a, b), b),$$

$(a, b)^2 \in F_3$. Similarly

$$1 = ((a, b), b^2) = ((a, b), b)^2(((a, b), b), b) = (a, b)^{-4}(((a, b), b), b).$$

By induction,

$$(a, b)^{2^{n-2}} \in F_n;$$

hence in F/F_n,

$$(a, b)^{2^{n-2}} = 1.$$

By **(8)**, the F_n, $n = 1, 2, \ldots$, are all distinct and hence (A, B) in F/F_n is exactly of order 2^{n-1} and F/F_n is of order 2^n.

That this is not a freak case can be seen from Theorem 4 below. Since finite nilpotent groups are direct products of prime power groups, it is sufficient for $n = 4$ to discuss the case of $p = 2$.

THEOREM 4. *Let $A_i = \{a_i\}$, $1 \leqslant i \leqslant t$ be cyclic groups of order 2^{r_i}. Let $r_1 \leqslant r_2 \leqslant \ldots \leqslant r_t$. Let $F = \prod^t_{i=1} * A_i$. Let $G = F/F_4$. Then every element of G can be expressed uniquely in the form*

(25) $$a_1^{c_1}a_2^{c_2}\ldots a_t^{c_t}\prod_{i<j}(a_i, a_j)^{c_{ij}}(a_i^2, a_j)^{c_{ij}^{(2)}}(a_i, a_j^2)^{c_{ij}^{(2)}}\cdot$$

$$\prod_{i<j<k}((a_i, a_j), a_k)^{c_{ijk}}((a_j, a_k), a_i)^{c_{jki}}$$

where the c_i, c_{ij}, $c_{ij}^{(2)}$, c_{ijk}, c_{jki} are integers modulo

$$2^{r_i}, 2^{r_i+1}, 2^{r_i-1}, 2^{r_i}, 2^{r_i}$$

respectively while $c_{ij}^{(3)}$ are integers modulo 2^{r_i-1}, if $r_i = r_j$, and 2^{r_i} if $r_i \neq r_j$. In particular, (a_i, a_j) is of order 2^{r_i+1} for $i \neq j$.

Formulas for multiplying two elements of G are given below.

Proof. Let a, b, c be three of the a_i of orders n_a, n_b, n_c, respectively, $n_a \leqslant n_b \leqslant n_c$. By (16)

$$1 = (a^{n_a}, b) = (a, b)^{n_a}(a, b, a)^{\binom{n_a}{2}}, a, b \in G.$$

From the work of Magnus **(11)**, it follows that

$$1 = (a, b, a)^{n_a} = (a, b, b)^{n_a} = (a, b, c)^{n_a} = (b, c, a)^{n_a} \quad \text{in } G.$$

Since (G, G) is Abelian, and $\binom{n_a}{2} \equiv n_a/2 \pmod{n_a}$

(26) $$(a, b)^{2n_a} = 1$$

and

(27) $$(a, b)^{-n_a} = (a, b)^{n_a} = ((a, b), a)^{\frac{1}{2}n_a}.$$

If $n_a = n_b$, the same reasoning gives

$$(a, b)^{n_a} = ((a, b), b)^{\frac{1}{2}n_a}.$$

However, if $n_a < n_b$, all that can be said is $((a, b), b)^{n_a} = 1$. In view of (26) and (27), computing a multiplication table using a representation such as (18) would be somewhat complicated; to avoid this difficulty, note that in G

$$(28) \qquad \begin{aligned} (a^2, b) &= (a, b)^2((a, b), a) \\ (a, b^2) &= (a, b)^2((a, b), b) \end{aligned}$$

and hence $\{(a, b), ((a, b), a), ((a, b), b)\} = \{(a, b), (a, b^2), (a^2, b)\}$. Now, using (27) and the fact that (G, G) is Abelian,

$$(a^2, b)^{\frac{1}{2}n_a} = (a, b)^{n_a}((a, b), a)^{\frac{1}{2}n_a} = 1.$$

If $n_a = n_b$, then $(a, b^2)^{\frac{1}{2}n_a} = 1$, while if $n_a < n_b$,

$$(a, b^2)^{n_a} = (a, b)^{2n_a}((a, b), b)^{n_a} = 1.$$

Hence every element of G can be expressed in the form of (25). If one multiplies two elements like (25), that is, let

$$\begin{aligned} c &= a_1^{c_1}a_2^{c_2}\cdots & (a_i, a_j)^{c_{ij}}\cdots & & ((a_i, a_j), a_k)^{c_{ijk}}\cdots \\ d &= a_1^{d_1}a_2^{d_2}\cdots & (a_i, a_j)^{d_{ij}}\cdots & & ((a_i, a_j), a_k)^{d_{ijk}}\cdots \\ e &= a_1^{e_1}a_2^{e_2}\cdots & (a_i, a_j)^{e_{ij}}\cdots & & ((a_i, a_j), a_k)^{e_{ijk}}\cdots \end{aligned}$$

with $e = c\cdot d$, then

$$e_i = c_i + d_i$$

$$e_{ij} = c_{ij} + d_{ij} - 2\alpha(c_{ij})d_i - 2\alpha(c_{ij})d_j - c_jd_i + 2c_j\binom{d_i}{2} + 2d_i\binom{c_j}{2} + 2\,c_jd_id_j$$

$$(29) \qquad e_{ij}^{(3)} = c_{ij}^{(3)} + d_{ij}^{(3)} + \alpha(c_{ij})d_i - c_j\binom{d_i}{2}$$

$$e_{ij}^{(3)} = c_{ij}^{(3)} + d_{ij}^{(3)} - d_i\binom{c_j}{2} + \alpha(c_{ij})d_j - c_jd_id_j$$

$$e_{ijk} = c_{ijk} + d_{ijk} + \alpha(c_{ik})d_j - c_kd_id_j - d_ic_jc_k + \alpha(c_{ij})d_k - c_jd_id_k$$

$$e_{jki} = c_{jki} + d_{jki} + \alpha(c_{jk})d_i + \alpha(c_{ik})d_j - c_kd_id_j$$

where

$$\alpha(c_{ij}) = c_{ij} + 2c_{ij}^{(2)} + 2c_{ij}^{(3)}.$$

Here there appear to be a few problems as to ambiguities, since, for example, d_i is an integer modulo 2^{r_i} and appears in the computation of e_{ij} which is an integer modulo 2^{r_i+1}. However, if d_i is replaced by $d_i + 2^{r_i}$, then

$$-c_jd_i + 2\,c_j\binom{d_i}{2} + 2d_i\binom{c_j}{2}$$

remains unchanged modulo 2^{r_i+1}. Similar reasoning applies to other cases of apparent ambiguity.

We can now proceed as in the proof of Theorem 1 and construct a group H made of

$$t + 3\binom{t}{2} + 2\binom{t}{3} - \text{tples}$$

with multiplication as indicated by (29). The verification of the group axioms is straightforward, but tedious.

It might be asked whether or not a modification of (18) could not be used instead of (29). There are several difficulties: in the case of $p = 2$, the e_{ij} are integers modulo 2^{r_i+1}, but c_j, d_i which appear in the formula for e_{ij} are integers modulo 2^{r_i} (assuming $r_i = r_j$). Similarly if $r_i = r_j$, e_{ijt} is an integer modulo 2^{r_i} and $\binom{d_i}{2}$ will cause difficulties, since it is not unambiguously defined modulo 2^{r_i}. If one decides to let c_{ij} be integers modulo 2^{r_i}, then the fact that

$$(a_i, a_j)^{2^{r_i}} = (a_i, a_j, a_i)^{2^{r_i-1}} \qquad \text{(see (27))}$$

means that the multiplication formulas would have to take into account in some way the fact that the order of (a_i, a_j) is 2^{r_i+1}. The author tried to think of a device to get around these difficulties, but was unable to do so.

If one attempts to carry out computations for the general case, with $p < n - 1$, then by using (5) and (6) one readily obtains Lemma 3 below. Since nilpotent groups of finite order are direct products of p-groups, we consider only the case of p-groups here.

LEMMA 3. *Let $A_1, \ldots, A_t$ be cyclic groups of order*

$$p^{a_1}, \ldots, p^{a_t}$$

respectively. Let a_i generate A_i. Let $F = \prod^t_{i=1} {}^* A_i$. Let $G = F/F_n$; let

$$v = (a_{i_1}, a_{i_2}, \ldots, a_{i_r}) \in G_r.$$

Let

$$\alpha = \min(\alpha_{i_1}, \ldots, \alpha_{i_r}).$$

Then

$$v^{p^\alpha} \in G_{r+(p-1)} \tag{30}$$

$$v^{p^{\alpha+j}} \in G_{r+(j+1)(p-1)} \qquad j = 0, 1, 2 \ldots.$$

If $w \in G_r$, then w can be substituted for v in (30).

Proof. The proof follows by induction ($r = n - 1, n - 2, \ldots,$) and uses (6) and (4).

Note that (20) is a special case of (30) with $w = a^i b^j$, $r = 1$, $p = 3$, $\alpha = 1$, $j = 0$, $n = 4$. Similarly, using group (23), one obtains another special case

of Lemma 3, with $w = ab$, $r = 1$, $p = 2$, $n = 3$, $\alpha = 1$, $j = 0$. This gives rise to the conjecture that these may be the best possible results in the following sense:

Conjecture. In the notation of Lemma 3, the order of v is $p^{\alpha+j}$, where j is the least integer such that

$$r + (j + 1)(p - 1) \geqslant n.$$

However, the author was unable to think of a way to prove that the order of v is exactly $p^{\alpha+j}$ and not something less, nor of a manageable method to solve the general case of $p < n - 1$.

References

1. O. N. Golovin, *Nilpotent products of groups*, Mat. Sbornik N.D., *27* (*69*) (1950), 427–454. Amer. Math. Soc. Translations, 2, *2* (1956), 89–115.
2. ——— *Metabelian products of groups*, Mat. Sbornik N.S., *28* (*70*) (1951), 431–444. Amer. Math. Soc. Translations, 2, *2* (1956), 117–132.
3. ——— *On the isomorphism of nilpotent decompositions of groups*, Mat. Sbornik N.S., *28* (*70*) (1951), 445–452. Amer. Math. Soc. Translations, 2, *2* (1956), 133–140.
4. K. W. Gruenberg, *Residual properties of infinite soluble groups*, Proc. London Math. Soc., Series 3, *7* (1957), 29–62.
5. Philip Hall, *A contribution to the theory of groups of prime-power order*, Proc. London Math. Soc., *36* (1934), 29–95.
6. ——— *Nilpotent groups*, Lecture Notes of Summer Seminar, Canadian Mathematical Congress (University of Alberta, August, 1957).
7. Marshall Hall, *A basis for free Lie rings and higher commutators in free groups*, Proc. Amer. Math. Soc., *1* (1950), 575–581.
8. A. Karass and D. Solitar, *On free products of groups*, Bull. Amer. Math. Soc., *63* (1957), 407
9. Friedrich Levi and B. L. van der Waerden, *Ueber eine Besondere Klasse von Gruppen*, Abhandlungen aus dem Hamburg Universität., *9* (1932), 154–158.
10. R C Lyndon, *On Burnside's problem, I.* Trans. Amer. Math. Soc., *77* (1954), 202–215.
11. W. Magnus, *Ueber Beziehungen zwischen höheren Kommutatoren*, J. Reine Angew. Math., *177* (1937), 105–115.
12. S. Moran, *Associative operations on groups* I. Proc. London Math. Soc., *6* (1956), 581–596.
13. I. N. Sanov, *Establishment of a connection between periodic groups with prime power periods and Lie rings.* Izvestiya Akad Nauk SSSR Ser Mat., *16* (1952), 23–58.
14. R. R. Struik, *Notes on a paper by sanov*, Proc. Amer. Math. Soc., *8* (1957), 638–641.
15. ——— *A note on prime power groups*, Can. Math. Bull., *3* (1960), 27–30.

University of British Columbia

TRACES OF MATRICES OF ZEROS AND ONES

H. J. RYSER

1. Introduction. This paper continues the study appearing in **(9)** and **(10)** of the combinatorial properties of a matrix A of m rows and n columns, all of whose entries are 0's and 1's. Let the sum of row i of A be denoted by r_i and let the sum of column j of A be denoted by s_j. We call $R = (r_1, \ldots, r_m)$ the *row sum vector* and $S = (s_1, \ldots, s_n)$ the *column sum vector* of A. The vectors R and S determine a class

$$\mathfrak{A} = \mathfrak{A}(R, S) \tag{1.1}$$

consisting of all $(0, 1)$-matrices of m rows and n columns, with row sum vector R and column sum vector S. The majorization concept yields simple necessary and sufficient conditions on R and S in order that the class $\mathfrak{A}$ be non-empty **(4; 9)**. Generalizations of this result and a critical survey of a wide variety of related problems are available in **(6)**.

Consider the 2 by 2 submatrices of A of the types

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

An *interchange* is a transformation of the elements of A which changes a minor of type A_1 into type A_2, or vice versa, and leaves all other elements of A unaltered. The interchange theorem (9) asserts that if A and A^* belong to A, then A is transformable into A^* by interchanges.

The *term rank* ρ of A is the order of the greatest minor of A with a non-zero term in its determinant expansion **(8)**. This integer equals the minimal number of rows and columns which contain collectively all of the non-zero elements of A **(7)**. Now let $\bar{\rho}$ be the maximal and $\underline{\rho}$ the minimal term rank for the matrices in $\mathfrak{A}$. The interchange theorem implies the existence of a matrix A in $\mathfrak{A}$ of term rank ρ **(9)**. Here ρ is an arbitrary integer in the interval

$$\underline{\rho} \leqslant \rho \leqslant \bar{\rho}. \tag{1.2}$$

Let $\delta_i = (1, \ldots, 1, 0, \ldots, 0)$ be a vector of n components, with 1's in the first r_i positions and 0's elsewhere. The matrix

$$\bar{A} = \begin{bmatrix} \delta_1 \\ \ldots \\ \delta_m \end{bmatrix} \tag{1.3}$$

Received April 7, 1959. This work was sponsored in part by the Office of Ordnance Research.

is called *maximal*, and $\bar{A}$ is the *maximal form* of A. Suppose that the components of R and S are positive. Define

$$R' = (r_1 - 1, \ldots, r_m - 1)$$

and let $\bar{A}'$ be the maximal matrix of m rows and n columns with row sum vector R'. Let the column sum vector of $\bar{A}'$ equal

$$\bar{S}' = (\bar{s}_1', \ldots, \bar{s}_n').$$

Renumber the subscripts of the column sum vector $S = (s_1, \ldots, s_n)$ so that

$$s_1 \geqslant \ldots \geqslant s_n.$$

Define

$$s_i' = s_i - 1 \qquad (i = 1, \ldots, n),$$
$$\bar{s}_0' = s_0' = 0,$$

and let

$$M = \max\left(\sum_{i=0}^{k} (s_i' - \bar{s}_i')\right) \qquad (k = 0, \ldots, n). \tag{1.4}$$

One may prove **(10)**

$$\bar{\rho} = m - M. \tag{1.5}$$

A simple formula for $\bar{\rho}$ analogous to (1.5) for $\bar{\rho}$ does not appear to exist. However, Haber in a forthcoming paper obtains an algorithm that yields an effective procedure for the determination of $\bar{\rho}$ (5).

Throughout the discussion we suppose that $\mathfrak{A}$ is non-empty and that $R = (r_1, \ldots, r_m)$ and $S = (s_1, \ldots, s_n)$ satisfy

$$r_1 \geqslant \ldots \geqslant r_m > 0, \tag{1.6}$$

$$s_1 \geqslant \ldots \geqslant s_n > 0. \tag{1.7}$$

We call the above R and S and the associated $\mathfrak{A}$ of (0, 1)-matrices *normalized*. Term rank is invariant under permutations of rows and columns. Thus normalization does not restrict this concept. Indeed, formula (1.5) for $\bar{\rho}$ actually requires a normalized S.

For $A = [a_{rs}]$ in $\mathfrak{A}$ we define the *trace* of A by

$$\operatorname{tr}(A) = \sum_{i=1}^{\epsilon} a_{ii}, \tag{1.8}$$

where

$$\epsilon = \min(m, n). \tag{1.9}$$

Fulkerson has recently investigated feasibility conditions for the existence of a (0, 1)-matrix of order n with specified row and column sums and 0 trace **(3)**. He utilizes the theory of network flows **(1; 2; 4)** and obtains an especially simple criterion for the case in which $\mathfrak{A}$ is normalized **(3)**. Let $\bar{\sigma}$ be the maximal

and $\bar{\sigma}$ the minimal trace for the matrices in the normalized $\mathfrak{A}$. In the present paper we develop a trace theory for $\bar{\sigma}$ and $\tilde{\sigma}$ analogous to the term rank theory for $\bar{\rho}$ and $\tilde{\rho}$. The requirement of positive components on R and S is without loss of generality. The ordering of the components in accordance with (1.6) and (1.7) does impose a restriction. But this is necessary in order to obtain conclusions of the uncomplicated type to be described.

Let A be in the normalized $\mathfrak{A}$ and write

$$A = \begin{bmatrix} W & X \\ Y & Z \end{bmatrix}, \tag{1.10}$$

where W is of size e by f $(0 \leqslant e \leqslant m; 0 \leqslant f \leqslant n)$. For an arbitrary $(0, 1)$-matrix Q, let $N_0(Q)$ denote the number of 0's in Q and $N_1(Q)$ the number of 1's in Q. Let

$$t_{ef} = N_0(W) + N_1(Z) \qquad (e = 0, \ldots, m; f = 0, \ldots, n) \tag{1.11}$$

and define

$$T = [t_{ef}] \qquad (e = 0, \ldots, m;\ f = 0, \ldots, n). \tag{1.12}$$

T is called the *structure matrix of the class* $\mathfrak{A}$. Its elementary properties are developed in § 2. Section 3 yields explicit formulae for $\bar{\sigma}$ and $\tilde{\sigma}$ in terms of the entries of T:

$$\bar{\sigma} = \min_{e,f} \{t_{ef} + \max(e, f)\} \tag{1.13}$$

$$\tilde{\sigma} = \max_{e,f} \{\min(e, f) - t_{ef}\} \qquad (e = 0, \ldots, m; f = 0, \ldots, n). \tag{1.14}$$

Matrices with an unusually simple block decomposition are shown to exist for the case of maximal and minimal trace, and these matrices play an essential role in the derivations of (1.13) and (1.14). Section 4 stresses similarities and differences in the behaviour of trace and term rank. The paper concludes with the determination of the domain of intermediate values for the traces σ of the matrices in $\mathfrak{A}$. This usually consists of all integers in the interval

$$\bar{\sigma} \leqslant \sigma \leqslant \tilde{\sigma}. \tag{1.15}$$

But certain classes $\mathfrak{A}$ exclude $\bar{\sigma} + 1$ and others exclude $\tilde{\sigma} - 1$.

2. The structure matrix. Let A belong to the normalized class $\mathfrak{A} = \mathfrak{A}(R, S)$ described in § 1 and write

$$A = \begin{bmatrix} W & X \\ Y & Z \end{bmatrix}, \tag{2.1}$$

where W is of size e by f $(0 \leqslant e \leqslant m; 0 \leqslant f \leqslant n)$. Let

$$T = [t_{ef}] \qquad (e = 0, \ldots, m;\ f = 0, \ldots, n) \tag{2.2}$$

denote the structure matrix of $\mathfrak{A}$. This means that

$$t_{ef} = N_0(W) + N_1(Z) \qquad (e = 0, \ldots, m;\ f = 0, \ldots, n). \tag{2.3}$$

where $N_0(W)$ denotes the number of 0's in W and $N_1(Z)$ the number of 1's in Z. It follows at once from (2.3) that

$$t_{ef} = ef + (r_{e+1} + \ldots + r_m) - (s_1 + \ldots + s_f) \qquad (e = 0, \ldots, m;\ f = 0, \ldots, n). \tag{2.4}$$

Thus the structure matrix is independent of the particular choice of A in $\mathfrak{A}$. Note that if

$$\tau = N_1(A) = r_1 + \ldots + r_m, \tag{2.5}$$

then the first row and column of T are given by

$$\begin{aligned} t_{0f} &= \tau - (s_1 + \ldots + s_f), \\ t_{e0} &= \tau - (r_1 + \ldots + r_e). \end{aligned} \tag{2.6}$$

The structure matrix has a number of interesting properties that give insight into the combinatorial behaviour of $\mathfrak{A}$. Its entries are, of course, non-negative integers and its size is $m + 1$ by $n + 1$. For notational convenience we number the rows of a matrix of these dimensions from 0 through m and its columns from 0 through n. Let E_k be the triangular matrix of order $k + 1$, with 1's on and below the main diagonal and 0's elsewhere. Let E_k^T denote the transpose of E_k, and number the rows and columns of E_k and E_k^T from 0 through k. Let S be the m by n matrix of 1's. Then

$$E_m \begin{bmatrix} \tau & -s_1 & \ldots & -s_n \\ -r_1 & & & \\ \vdots & & S & \\ -r_m & & & \end{bmatrix} E_n^T = T. \tag{2.7}$$

For the eth row vector of the product of the first two matrices on the left side of equation (2.7) is

$$(r_{e+1} + \ldots + r_m, -s_1 + e, \ldots, -s_n + e).$$

If this row is multiplied by the fth column of E_n^T, then we obtain

$$ef + (r_{e+1} + \ldots + r_m) - (s_1 + \ldots + s_f). \tag{2.8}$$

But by (2.4) this is t_{ef}.

If in (2.4) we replace f by $f + 1$, then

$$t_{e,f+1} = ef + e + (r_{e+1} + \ldots + r_m) - (s_1 + \ldots + s_{f+1}) \qquad (e = 0, \ldots, m;\ f = 0, \ldots, n - 1). \tag{2.9}$$

By (2.4) and (2.9),

$$t_{e,f+1} = t_{ef} + e - s_{f+1} \qquad (e = 0, \ldots, m;\ f = 0, \ldots, n-1). \tag{2.10}$$

Similarly, we may deduce

$$t_{e+1,f} = t_{ef} + f - r_{e+1} \qquad (e = 0, \ldots, m-1;\ f = 0, \ldots, n). \tag{2.11}$$

The recursions (2.10) and (2.11) are useful in constructing T from a given R and S.

From (2.10) we see that the eth row of T may be written in the form

$$(t_{e0}, t_{e0} + e - s_1, t_{e1} + e - s_2, \ldots, t_{e,n-1} + e - s_n). \tag{2.12}$$

S is normalized so that by (2.12), if $e \leqslant s_n$, then

$$t_{e0} \geqslant t_{e1} \geqslant \ldots \geqslant t_{en}, \tag{2.13}$$

and if $e \geqslant s_1$, then

$$t_{e0} \leqslant t_{e1} \leqslant \ldots \leqslant t_{en}. \tag{2.14}$$

On the other hand, if $s_n < e < s_1$, then there must exist an integer $f (0 < f < n)$ such that

$$t_{e0} \geqslant t_{e1} \geqslant \ldots \geqslant t_{ef} \leqslant t_{e,f+1} \leqslant \ldots \leqslant t_{en}. \tag{2.15}$$

The columns of T have an analogous monotonic behaviour.

The following numerical example affords a simple illustration of the preceding remarks:

$$R = (4, 3, 2, 2, 1), \qquad S = (4, 4, 2, 1, 1),$$

$$T = \begin{bmatrix} 12 & 8 & 4 & 2 & 1 & 0 \\ 8 & 5 & 2 & 1 & 1 & 1 \\ 5 & 3 & 1 & 1 & 2 & 3 \\ 3 & 2 & 1 & 2 & 4 & 6 \\ 1 & 1 & 1 & 3 & 6 & 9 \\ 0 & 1 & 2 & 5 & 9 & 13 \end{bmatrix}.$$

3. Maximal and minimal traces. Let $\bar{\sigma}$ be the maximal and $\underline{\sigma}$ the minimal trace for the matrices in the normalized class $\mathfrak{A}$. In this section we develop simple block decompositions for the matrices of maximal and minimal trace and use these decompositions to derive (1.13) and (1.14). We begin with an elementary property of the trace function for the class $\mathfrak{A}$.

THEOREM 3.1. *Suppose the normalized $\mathfrak{A}$ contains a matrix of trace σ. Then there exists an $A = [a_{rs}]$ in $\mathfrak{A}$ of trace σ with the 1's in the initial positions on the main diagonal*

$$a_{11} = \ldots = a_{\epsilon\epsilon} = 1, \tag{3.1}$$
$$a_{\epsilon+1,\epsilon+1} = \ldots = a_{ss} = 0 \qquad (\epsilon = \min(m, n)).$$

For suppose that we have an A in $\mathfrak{A}$ of trace σ with

$$a_{11} = \ldots = a_{\epsilon-1,\epsilon-1} = 1 \qquad (\epsilon - 1 < \sigma), \tag{3.2}$$
$$a_{\epsilon\epsilon} = 0.$$

It suffices to show that it is possible to transform A by interchanges into a matrix of trace σ with the ϵ leading diagonal elements equal to 1. Now there must exist an integer $t > \sigma$ such that $a_{tt} = 1$. Suppose that

$$a_{\epsilon t} = a_{t\epsilon} = 0. \tag{3.3}$$

Since $r_\epsilon \geqslant r_t$ and $s_\epsilon \geqslant s_t$, there exist integers u and v such that $a_{u\epsilon} = 1$, $a_{ut} = 0$ and $a_{\epsilon v} = 1$, $a_{tv} = 0$. We apply an interchange involving positions (ϵ, t), (ϵ, v), (t, v), (t, t), and follow this by an interchange involving positions (ϵ, ϵ), (ϵ, t), (u, t), (u, ϵ). This gives a 0 in the (t, t) position and a 1 in the (ϵ, ϵ) position. The other diagonal elements are not altered. The remaining cases

$$a_{\epsilon t} = a_{t\epsilon} = 1, \tag{3.4}$$

$$a_{\epsilon t} = 1, \quad a_{t\epsilon} = 0, \tag{3.5}$$

$$a_{\epsilon t} = 0, \quad a_{t\epsilon} = 1, \tag{3.6}$$

are disposed of by similar arguments.

We turn now to a study of the maximal trace $\bar{\sigma}$ for matrices in the class $\mathfrak{A}$.

THEOREM 3.2. *Let $\bar{\sigma} \neq \min(m, n)$. Then there exists a matrix $A_{\bar{\sigma}}$ of trace $\bar{\sigma}$ in the normalized $\mathfrak{A}$ of the form*

$$A_{\bar{\sigma}} = \begin{bmatrix} S & * & * \\ * & \bar{0} & 0 \\ * & 0 & 0 \end{bmatrix}. \tag{3.7}$$

Here S is a matrix of 1's *of specified size e by f $(0 < e \leqslant \bar{\sigma};\ 0 < f \leqslant \bar{\sigma})$. The matrix $\bar{0}$ is of size g by h and has* 1's *in the main diagonal positions of $A_{\bar{\sigma}}$ and* 0's *in all other positions. Moreover,*

$$e + g = f + h = \bar{\sigma}. \tag{3.8}$$

The 0's *denote zero matrices.*

Consider a matrix A in $\mathfrak{A}$ with the $\bar{\sigma}$ 1's in the initial positions on the main diagonal. We have $\bar{\sigma} > 0$. The block in the lower right corner of size $m - \bar{\sigma}$ by $n - \bar{\sigma}$ must be a zero block. For the row and column sum vectors are normalized and if the block contained a 1, then a suitable interchange would increase $\bar{\sigma}$. Now the matrix A may be selected to be of the following form:

$$A = \begin{bmatrix} S_1 & * & R_1 \\ * & * & 0_1 \\ C_1 & 0_2 & 0 \end{bmatrix}. \tag{3.9}$$

Here 0 is the zero block of size $m - \bar{\sigma}$ by $n - \bar{\sigma}$, 0_1 and 0_2 are zero blocks, R_1 has at least one 1 in each row, and C_1 has at least one 1 in each column. For let us consider two vectors X_1 and X_2 from among the first $\bar{\sigma}$ rows of A and let the entries of these vectors total x_1 and x_2, resepctively. Let X_1 have 0's in its last $n - \bar{\sigma}$ positions and let X_2 have a 1 in at least one of these positions. Suppose X_1 is above X_2 in A. If $x_1 > x_2$, then we may apply an interchange involving X_1 and X_2 that does not shift a 1 on the main diagonal and places a 1 in one of the last $n - \bar{\sigma}$ positions of X_1. If $x_1 = x_2$, we may still apply the interchange and place a 1 in one of the last $n - \bar{\sigma}$ positions of X_1. However, in exceptional cases a single interchange may be available for this purpose and this may force a reduction in trace to $\bar{\sigma} - 1$. But if this is the case, then a second interchange confined to the first $\bar{\sigma}$ columns restores trace $\bar{\sigma}$. This procedure yields the blocks R_1 and 0_1 of (3.9). Next we work on the columns and in the same way. Again a single interchange may be available and force a reduction in trace to $\bar{\sigma} - 1$. But under these circumstances there is always available a second interchange confined to the first $\bar{\sigma}$ rows and columns that regains trace $\bar{\sigma}$. This is the case since otherwise an interchange exists that restores the 1 to the main diagonal and places a 1 in the block 0, contradicting the maximality of $\bar{\sigma}$. This gives us a matrix of the form (3.9).

Now let S_1 of (3.9) be of size $\bar{e}$ by $\bar{f}$. S_1 must be a block of 1's, for otherwise we could increase $\bar{\sigma}$. An A of the form (3.9) with fixed $\bar{e}$ and $\bar{f}$ we call *reduced*. The $\bar{\sigma}$ 1's on its main diagonal we call *essential* 1's. All other 1's are called *unessential*. Without loss of generality we may assume $\bar{e} \leqslant \bar{f}$.

We now consider a reduced A^* in $\mathfrak{A}$ of the form

$$\mathrm{A}^* = \left[\begin{array}{c|c|c|c} S_1 & S_2 & Y & R_1 \\ \hline \multicolumn{2}{c|}{X} & Z & 0_1 \\ \hline C_1 & \multicolumn{2}{c|}{0_2} & 0 \end{array}\right]. \tag{3.10}$$

Here S_2 is a block of 1's of size $\bar{e}$ by $f^* - \bar{f}$ with $f^* - \bar{f}$ maximal in A^*. Among all reduced A in $\mathfrak{A}$ we select that A^* with its corresponding $f^* - \bar{f}$ minimal. We must allow the case $f^* - \bar{f} = 0$. But if the minimal $f^* - \bar{f} > 0$, then S_2 is a block of 1's that appears in all of the reduced A in $\mathfrak{A}$. If Y is not present, then (3.7) holds with $h = 0$. Suppose then that Y is present. Then our A^* has a 0 in the first column of Y. If the block Z contains no unessential 1, then (3.7) holds with $f = f^*$. Suppose, therefore, that an unessential 1 appears in the (s, t) position of Z, where s is maximal in Z. If $t = 1$, then there is a 0 in column t of Y. Suppose that $t \neq 1$ and let the $(s, 1)$ position of Z contain

a 0 or an essential 1. If column t of Y contains only 1's, then we may perform an interchange using unessential 1's and the 0 in column 1 of Y to obtain a 0 in column t of Y. We henceforth require A^* to have a 0 in column t (but no longer column 1) of Y.

The preceding remarks imply that the entries in row s of X must be 1's. For if this is not the case then a single interchange gives a reduced A^* with a 0 in S_2, contradicting $f^* - \bar{f}$ minimal, or else two interchanges place a 1 in 0, contradicting $\bar{\sigma}$ maximal. Suppose that X has a 0 present in its (u, v) position, where $u < s$. Then we may apply an interchange involving this 0 and the 1 in the (s, v) position of X. If this interchange does not involve an essential 1, then a second interchange involving the unessential 1 in the (s, t) position of Z introduces a 0 into S_1 or S_2. This leads us to the same contradiction as before. Suppose then that the interchange involving the 0 in the (u, v) position and the 1 in the (s, v) position of X does involve an essential 1. Consider the case $s \leqslant f^* - \bar{e}$. Then a second interchange involving the unessential 1 in the (s, t) position of Z regains trace $\bar{\sigma}$ and introduces a 0 into S_1 or S_2. This is again a contradiction. If $s > f^* - \bar{e}$, then the second interchange involving the unessential 1 in the (s, t) position of Z introduces a 0 into S_1 or S_2. However, the trace of the matrix upon completion of this interchange remains at $\bar{\sigma} - 1$. But then we may apply a third interchange involving rows u and s of Z and regain trace $\bar{\sigma}$. Thus in all cases there is no 0 present in the (u, v) position of X, where $u < s$. This gives a matrix of the form (3.7) and completes the proof.

THEOREM 3.3. *The maximal trace $\bar{\sigma}$ for the matrices in the normalized $\mathfrak{A}$ is given by*

$$\bar{\sigma} = \min_{e,f} \{t_{ef} + \max(e, f)\} \qquad (e = 0, \ldots, m; f = 0, \ldots, n). \tag{3.11}$$

Let A be a matrix in $\mathfrak{A}$ of trace $\bar{\sigma}$ with the $\bar{\sigma}$ 1's in the initial positions on the main diagonal. Let A be subdivided into the four blocks W, X, Y, Z of (2.1) with W of size e by f. Now for the matrix A under consideration it is clear that

$$N_1(Z) \geqslant \bar{\sigma} - \max(e, f) \qquad (e = 0, \ldots, m; f = 0, \ldots, n). \tag{3.12}$$

But $N_0(W) \geqslant 0$ so that

$$t_{ef} + \max(e, f) = N_0(W) + N_1(Z) + \max(e, f) \geqslant \bar{\sigma} \qquad (e = 0, \ldots, m; f = 0, \ldots, n). \tag{3.13}$$

Suppose that $\bar{\sigma} \neq \min(m, n)$. Then we may specialize our A to the $A_{\bar{\sigma}}$ of Theorem 3.2. The submatrix S of $A_{\bar{\sigma}}$ is of size e by f. We may set $W = S$ and obtain $N_0(W) = 0$ and $N_1(Z) = \bar{\sigma} - \max(e, f)$. Thus if $\bar{\sigma} \neq \min(m, n)$,

equality is attained in (3.13) for the dimension numbers e and f of the submatrix S of $A_{\tilde{\sigma}}$. If $\tilde{\sigma} = m$, equality is attained in (3.13) for $f = 0$ and $e = m$. If $\tilde{\sigma} = n$, equality is attained in (3.13) for $e = 0$ and $f = n$. This proves Theorem 3.3.

We consider next the minimal trace $\tilde{\sigma}$ for the matrices in $\mathfrak{A}$.

THEOREM 3.4. *Let the matrices in the normalized $\mathfrak{A}$ have precisely u rows and v columns composed entirely of* 1's *and let $\tilde{\sigma} \neq \max(u, v)$. Then there exists a matrix $A_{\tilde{\sigma}}$ of trace $\tilde{\sigma}$ in $\mathfrak{A}$ of the form*

$$A_{\tilde{\sigma}} = \begin{bmatrix} S & S_1 & * \\ S_2 & \bar{S} & * \\ * & * & 0 \end{bmatrix}. \tag{3.14}$$

Here S is a matrix of 1's *of order $\tilde{\sigma}$. S_1 of size $\tilde{\sigma}$ by s and S_2 of size t by $\tilde{\sigma}$ are matrices of* 1's. *$\bar{S}$ is a matrix with* 0's *in the main diagonal positions of $A_{\tilde{\sigma}}$ and* 1's *in all other positions.* 0 *is a zero matrix. (The cases $s = 0$ and $t = 0$ are not excluded.)*

Let A be a matrix in $\mathfrak{A}$. If A is not square, then add zero rows at the bottom or zero columns at the right and obtain a square matrix $\bar{A}$ of order $\max(m, n)$. In $\bar{A}$ replace the 1's by 0's and the 0's by 1's. This yields a matrix $\bar{C}$ called the *complement* of $\bar{A}$. The matrix $\bar{C}$ determines a class $\mathfrak{C}$. Let $\bar{C}$ be a matrix in $\mathfrak{C}$ of maximal trace $\tilde{\sigma}_c$. Evidently

$$\tilde{\sigma} = \max(m, n) - \tilde{\sigma}_c. \tag{3.15}$$

The matrix $\bar{C}$ has row sums and column sums in ascending order. Moreover, $\tilde{\sigma}_c \neq \max(m, n) - \max(u, v)$, for otherwise $\tilde{\sigma} = \max(u, v)$. We now apply Theorem 3.2 to the block in the lower right corner of $\bar{C}$ of size $\max(m, n) - u$ by $\max(m, n) - v$. This tells us that $\bar{C}$ may be written in the form

$$\bar{C} = \left[\begin{array}{c|c|c} 0 & 0 & * \\ \hline 0 & \begin{array}{c|c} 0 & 0 \\ \hline 0 & \bar{0} \end{array} & * \\ \hline * & * & S \end{array}\right]. \tag{3.16}$$

The 0's denote zero blocks. The 0 in the upper left corner of $\bar{C}$ is of size u by v. S is a block of 1's of size $\bar{e}$ by $\bar{f}$. $\bar{0}$ of size $\bar{g}$ by $\bar{h}$ has 1's in the main diagonal positions of $\bar{C}$ and 0's in all other positions. Moreover,

$$\bar{e} + \bar{g} = \bar{f} + \bar{h} = \tilde{\sigma}_c. \tag{3.17}$$

Now take the complement of $\bar{C}$ and delete all zero rows or columns. This yields a matrix $A_{\tilde{\sigma}}$ of the type described in the theorem.

THEOREM 3.5. *The minimal trace $\bar{\sigma}$ for the matrices in the normalized $\mathfrak{A}$ is given by*

$$\bar{\sigma} = \max_{e,f} \{\min(e,f) - t_{ef}\} \tag{3.18}$$
$$(e = 0, \ldots, m; f = 0, \ldots, n).$$

Let A be a matrix in $\mathfrak{A}$ of trace $\bar{\sigma}$ with the $\bar{\sigma}$ 1's in the initial positions on the main diagonal. Let A be subdivided into the four blocks W, X, Y, Z of (2.1) with W of size e by f. Now for the matrix A under consideration, it is clear that

$$N_0(W) \geqslant \min(e,f) - \bar{\sigma} \tag{3.19}$$
$$(e = 0, \ldots, m; f = 0, \ldots, n).$$

But $N_1(Z) \geqslant 0$ so that

$$\min(e,f) - t_{ef} = \min(e,f) - (N_0(W) + N_1(Z)) \leqslant \bar{\sigma} \tag{3.20}$$
$$(e = 0, \ldots, m; f = 0, \ldots, n).$$

Suppose that $\bar{\sigma} \neq \max(u,v)$, where the matrices in $\mathfrak{A}$ have precisely u rows and v columns composed entirely of 1's. Then we may specialize our A to the $A_{\bar{\sigma}}$ of Theorem 3.4. The matrix $A_{\bar{\sigma}}$ yields a W, X, Y, Z block subdivision with W of size e by f for which $N_0(W) = \min(e,f) - \bar{\sigma}$ and $N_1(Z) = 0$. Thus if $\bar{\sigma} \neq \max(u,v)$, then there exists an e and an f for which equality is attained in (3.20). If $\bar{\sigma} = u$, equality is attained in (3.20) for $e = \bar{\sigma}$ and $f = n$. If $\bar{\sigma} = v$, equality is attained in (3.20) for $e = m$ and $f = \bar{\sigma}$.

4. Trace and term rank. Let A belong to the normalized class $\mathfrak{A}$, and let $\bar{\rho}$ be the maximal term rank for the matrices in $\mathfrak{A}$. The integer $\bar{\rho}$ is given explicitly by (1.5). We derive a second formula for $\bar{\rho}$ analogous to (3.11) for $\bar{\sigma}$.

THEOREM 4.1. *The maximal term rank $\bar{\rho}$ for the matrices in the normalized $\mathfrak{A}$ is given by*

$$\bar{\rho} = \min_{e,f} \{t_{ef} + (e+f)\} \tag{4.1}$$
$$(e = 0, \ldots, m; f = 0, \ldots, n).$$

Let A be in the normalized $\mathfrak{A}$ and of maximal term rank $\bar{\rho}$. Let A be subdivided into the four blocks W, X, Y, Z of (2.1) with W of size e by f. Now the term rank of a matrix equals the minimal number of rows and columns which contain collectively all of the non-zero elements of the matrix. Hence for the matrix A under consideration, it is clear that

$$N_1(Z) + (e+f) \geqslant \bar{\rho}. \tag{4.2}$$

But $N_0(W) \geqslant 0$ so that

$$t_{ef} + (e+f) = N_0(W) + N_1(Z) + (e+f) \geqslant \bar{\rho}. \tag{4.3}$$

Suppose that $\bar{\rho} < \min(m, n)$. Then by Theorem 3.2 of (10) we may specialize our A to a matrix $A_{\bar{\rho}}$ of term rank $\bar{\rho}$ with a W of size e by f for which $N_0(W) = 0$ and $N_1(Z) = \bar{\rho} - (e + f)$. Thus if $\bar{\rho} \neq \min(m, n)$, equality is attained in (4.3) for the dimension numbers e and f of the submatrix W of $A_{\bar{\rho}}$. If $\bar{\rho} = m$, equality is attained in (4.3) for $f = 0$ and $e = m$. If $\bar{\rho} = n$, equality is attained in (4.3) for $e = 0$ and $f = n$. This establishes (4.1).

Let A belong to the normalized class $\mathfrak{A}$. An element $a_{rs} = 1$ of A is an *invariant* 1 provided that no sequence of interchanges applied to A replaces $a_{rs} = 1$ by 0 (10). If $a_{rs} = 1$ is an invariant 1 of A, then the entries in the (r, s) position of all of the matrices in $\mathfrak{A}$ must be invariant 1's. Thus all or none of the matrices in $\mathfrak{A}$ contains an invariant 1, and we say $\mathfrak{A}$ is with or without an invariant 1. The normalized class $\mathfrak{A}$ is with an invariant 1 if and only if the matrices in $\mathfrak{A}$ are of the form

$$A = \begin{bmatrix} S & * \\ * & 0 \end{bmatrix}. \tag{4.4}$$

Here S is a matrix of 1's of size e by f $(0 < e < m; 0 < f \leqslant n)$ and 0 is a zero block (10). Now the entries of the structure matrix T of $\mathfrak{A}$ are non-negative integers. Moreover,

$$t_{ef} > 0 \qquad (e = 1, \ldots, m;\ f = 1, \ldots, n) \tag{4.5}$$

if and only if $\mathfrak{A}$ is without an invariant 1. Indeed, each

$$t_{ef} = 0 \qquad (e, f > 0) \tag{4.6}$$

yields the dimension numbers e and f for a block decomposition of the type displayed in (4.4).

Let $\bar{\rho}$ be the maximal and $\tilde{\rho}$ the minimal term rank for the matrices in $\mathfrak{A}$. If $\bar{\rho} < \min(m, n)$ and if $\mathfrak{A}$ is without an invariant 1, then $\tilde{\rho} < \bar{\rho}$ (10). But important classes do exist with $\tilde{\rho} = \bar{\rho}$, for example, the class of all $(0, 1)$-matrices of order $m = n$ with exactly k 1's in each row and column. An unsettled problem asks for a neat classification of all $\mathfrak{A}$ with $\tilde{\rho} = \bar{\rho}$. The corresponding problem for traces in a normalized $\mathfrak{A}$ has an easy solution. For let A in $\mathfrak{A}$ be of trace $\tilde{\sigma}$ with 1's in the initial positions on the main diagonal. Then if $\tilde{\sigma} = \bar{\sigma}$, it follows readily that

$$A = \begin{bmatrix} S & * \\ * & 0 \end{bmatrix}. \tag{4.7}$$

Here S is the matrix of 1's of order $\tilde{\sigma}$ and 0 is a zero block. Thus a normalized class $\mathfrak{A}$ has $\tilde{\sigma} = \bar{\sigma}$ if and only if its structure matrix contains a zero on the main diagonal.

A single interchange alters the term rank of a matrix by at most 1. It follows from this and the interchange theorem that there exists an A in $\mathfrak{A}$ of term rank ρ, where ρ is an arbitrary integer such that

$$\tilde{\rho} \leqslant \rho \leqslant \bar{\rho}. \tag{4.8}$$

However, a single interchange may alter the trace σ of a matrix in $\mathfrak{A}$ by 2. This causes a complication in finding the domain of intermediate values for σ

$$\bar{\sigma} \leqslant \sigma \leqslant \tilde{\sigma}. \tag{4.9}$$

The problem of intermediate values is settled by the following theorem.

THEOREM 4.2. *The traces of the matrices in the normalized $\mathfrak{A}$ take on all integral values in the interval $\bar{\sigma} \leqslant \sigma \leqslant \tilde{\sigma}$ unless $\mathfrak{A}$ contains a matrix of the form*

$$A = \begin{bmatrix} S & S^* & * \\ S^{*T} & I_e & 0 \\ * & 0 & 0 \end{bmatrix}. \tag{4.10}$$

Here S is a matrix of 1's *of order e, S^* is a rectangular matrix of* 1's, *S^{*T} is the transpose of S^*, I_e is the identity matrix or the complement of this matrix, and the* 0's *are zero matrices. The order of I_e is g with $g \geqslant 2$. (The cases $e = 0$, $e + g = m$, and $e + g = n$ are not excluded.)*

Two matrices in $\mathfrak{A}$ are transformable into each other by interchanges, and a single interchange applied to a matrix in $\mathfrak{A}$ may alter its trace by at most 2. Consecutive traces of mtarices in $\mathfrak{A}$ may differ by at most 2. Suppose then that σ and $\sigma - 2$ but not $\sigma - 1$ appear as the traces of matrices in $\mathfrak{A}$. Then there exists an A_σ in $\mathfrak{A}$ of trace σ with a principal minor of order 2 that is the identity.

Thus there exists an A_σ in $\mathfrak{A}$ with a principal minor of order g

$$M = [m_{ij}] \qquad (i, j = 1, \ldots, g) \tag{4.11}$$

composed of consecutive rows and columns of A_σ and such that

$$m_{11} = m_{gg} = 1, m_{1g} = m_{g1} = 0. \tag{4.12}$$

We let g be maximal among all matrices in $\mathfrak{A}$ of trace σ and write

$$A_\sigma = \begin{bmatrix} \bar{A} & \bar{B} & D \\ \bar{C} & M & \bar{F} \\ E & \bar{G} & \bar{H} \end{bmatrix}. \tag{4.13}$$

The first row of A_σ passing through M must have the same sum as the last row of A_σ passing through M, for otherwise an interchange yields a trace of $\sigma - 1$. But $\mathfrak{A}$ is normalized so all rows of A_σ passing through M have the same sum. Similar remarks hold for the columns of A_σ passing through M.

Throughout the discussion we designate the submatrices of A_σ in (4.13) by $\bar{A} = [\bar{a}_{ij}]$, $\bar{F} = [\bar{f}_{ij}]$, etc. Suppose that in $\bar{F}$ some $\bar{f}_{uv} = 1$. If $\bar{f}_{1v} = 0$, we may apply an interchange involving $\bar{f}_{uv} = 1$ and $\bar{f}_{1v} = 0$. This interchange cannot yield a trace of $\sigma - 1$. Nor can the interchange increase the trace to $\sigma + 1$, for then an interchange involving $m_{11} = m_{gg} = 1$ yieods a trace of $\sigma - 1$.

Hence if some $\bar{f}_{uv} = 1$, then there exists an A_σ of the form (4.13) with $\bar{f}_{1v} = 1$. We may now apply an interchange involving $\bar{f}_{1v} = 1$ and $m_{1g} = 0$. If the trace remains equal to σ, a second interchange involving $m_{11} = 1$ and $m_{g1} = 0$ yields a trace of $\sigma - 1$. Suppose then that the interchange involving $\bar{f}_{1v} = 1$ and $m_{1g} = 0$ yields a trace of $\sigma + 1$. Let the 1 introduced on the main diagonal of A_σ be in the (t, t) position of $\bar{H}$. Then $\bar{g}_{t1} = 1$, for otherwise an interchange yields a trace of $\sigma - 1$. But now by an interchange involving $m_{gg} = \bar{g}_{t1} = 1$, we regain trace σ and contradict the maximality of g. Hence $\bar{F} = 0$. A similar argument gives $\bar{G} = 0$. By the maximality of g, each $\bar{h}_{uu} = 0$. If some $\bar{h}_{uv} = 1$ with $u \neq v$, then an interchange involving $\bar{h}_{uv} = m_{11} = 1$ yields a trace of $\sigma - 1$. Hence $\bar{H} = 0$.

Suppose that some $\bar{c}_{uv} = 0$. The rows of A_σ passing through M have the same sum, so that if some $\bar{c}_{uv} = 0$, then there exists an A_σ of the form (4.13) with $\bar{c}_{1v} = 0$. But then $\bar{a}_{vv} = 1$, $\bar{b}_{v1} = 0$ and $\bar{b}_{vg} = \bar{c}_{gv} = 0$, for otherwise an interchange yields a trace of $\sigma - 1$. But this contradicts the maximality of g. Hence $\bar{C}$ is a matrix of 1's. Similarly, $\bar{B}$ is a matrix of 1's.

Suppose that some $\bar{a}_{uu} = 0$. Then an interchange involving $\bar{a}_{uu} = m_{1g} = 0$ yields a trace of $\sigma + 1$. Now apply an interchange involving $m_{11} = \bar{c}_{gu} = 1$. This regains trace σ and contradicts the maximality of g. Hence each $\bar{a}_{uu} = 1$. If some $\bar{a}_{uv} = 0$ with $u \neq v$, then an interchange involving $\bar{a}_{uv} = m_{1g} = 0$ retains trace σ. A second interchange yields a trace of $\sigma - 1$. Hence $\bar{A}$ is a matrix of 1's.

All row and column sums of $M = [m_{ij}]$ must be equal. Suppose there exist u and v such that

$$m_{uu} = 1, \qquad m_{vv} = 0 \qquad (1 < u, v < g). \tag{4.14}$$

If $m_{uv} = 0$ or if $m_{vu} = 1$, an interchange yields a trace of $\sigma - 1$. Hence M has trace g or trace 2. Suppose M has trace g. If $m_{1t} = 1$ with $t > 1$, then $m_{gt} = 1$. An interchange involving columns t and g followed by an interchange involving columns 1 and t yields a trace of $\sigma - 1$. Hence row 1 of M has sum 1 and $M = I$. Suppose M has trace 2. If $m_{1t} = 0$ with $t < g$, then an interchange involving columns 1 and t yields a trace of $\sigma - 1$. Hence row 1 of M has sum $g - 1$ and an interchange replaces M by the complement of I. This proves Theorem 4.2.

It is clear that Theorem 4.2 disposes of the problem of finding the domain of intermediate values for the traces σ of the matrices in $\mathfrak{A}$. For suppose that $\mathfrak{A}$ contains a matrix of the form (4.10) and that I_c is the identity matrix. Then $\bar{\sigma} - 1$ is the single value excluded from the integers in the interval $\bar{\sigma} \leqslant \sigma \leqslant \bar{\sigma}$. Suppose on the other hand that $\mathfrak{A}$ contains a matrix of the form (4.10) and that I_c is the complement of the identity. Then $\bar{\sigma} + 1$ is the single value excluded from the integers in the interval $\bar{\sigma} \leqslant \sigma \leqslant \bar{\sigma}$.

References

1. L. R. Ford, Jr., and D. R. Fulkerson, *A simple algorithm for finding maximal network flows and an application to the Hitchcock problem*, Can. J. Math., *9* (1957), 210–218.
2. D. R. Fulkerson, *A network-flow feasibility theorem and combinatorial applications*, Can. J. Math., *11* (1959), 440–451.
3. ——— *Zero-one matrices with zero trace*, Rand Corporation publication P-1618.
4. David Gale, *A theorem on flows in networks*, Pac. J. Math., *7* (1957), 1073–1082.
5. R. M. Haber, *Term rank of* 0, 1 *matrices*, to appear in Ill. J. Math.
6. Alan J. Hoffman, *Some recent applications of the theory of linear inequalities to extermal combinatorial analysis*, to appear in the American Mathematical Society publication of the symposium on combinatorial designs and analysis.
7. Dénes König, *Theorie der endlichen und unendlichen Graphen* (New York, 1950).
8. Oystein Ore, *Graphs and matching theorems*, Duke Math. J., *22* (1955), 625–639.
9. H. J. Ryser, *Combinatorial properties of matrices of zeros and ones*, Can. J. Math., *9* (1957), 371–377.
10. ——— *The term rank of a matrix*, Can. J. Math., *10* (1958), 57–65.

The Ohio State University

A NEW TYPE OF CHARACTERISTIC SUBGROUP OF PRIME-POWER GROUPS

H. R. BRAHANA

1. Introduction. In a recent paper **(1)** the fifty-eight metabelian groups of order p^{11} that are generated by five elements and have all their elements of order p were determined and characterized in terms independent of any particular selection of the generating elements. In dealing with fifty-seven of these groups there was no occasion to distinguish between one odd prime and another, except that in exhibiting canonical forms it was necessary to select irreducible polynomials and these, of course, depended on p. The fifty-eighth group was described in two ways in terms that were independent of p, but the proof of uniqueness could not be made without taking into account properties of p. These properties distribute the primes into classes, and the properties are reflected in the groups of order p^{11} in characteristic subgroups some of which exist for one prime and not for another. It may be that examination of the groups of isomorphisms of some of the fifty-seven groups would produce characteristic subgroups for one p that would not exist for another, but the writer considers it doubtful. The doubt is made plausible by the fact that examination of some of the likeliest groups yielded no such subgroups, and by the belief that if a group is described and a canonical form obtained without making use of any special property of the prime then anything that is true for a group with one p will have an analogue for one with another. The fifty-eighth group has some characteristic subgroups pointed to by geometric differences appearing with different types of primes. It is believed this phenomenon of prime-power groups has not been brought to light before.

2. The groups. The following properties determine a group $\bar{G}$ of order p^{15} for every value of p:

1. Elements are all, except identity, of order p;
2. The group is metabelian;
3. Central and commutator subgroup coincide;[1]
4. The group has five generators.

The groups of order p^{11} are obtained by adding four conditions on commutators of this group, on commutators only because a condition that contained one of the generators explicitly would give a group that would not satisfy 3. Any four independent conditions on commutators will give a group of order p^{11}, in certain well-defined cases again violating 3.

Received June 15, 1959.

[1]Of course 3 includes 2.

To get the group we are seeking we require the four conditions to be such that G of order p^{11} satisfy:

5. G contains no abelian subgroup of order p^3;
6. G contains no subgroup of order p^{10} whose commutator subgroup is of order p^4.

G satisfying these conditions exists for every p and it is unique. In establishing the uniqueness it is necessary to make use of different arguments for different p's and this points to different characteristic subgroups of G.

Groups of order p^{11} satisfying $1, \ldots, 5$ also satisfy 6 or

6′. G contains one subgroup of order p^{10} whose commutator subgroup is of order p^4.

For purposes of comparison we shall consider this group too. Looked at geometrically the situation is not so mysterious.

Let U_1, U_2, U_3, U_4, U_5 be generators of the group $\bar{G}$ of order p^{15}; let c_{ij} be the commutator of U_i and U_j; and let $\bar{C}$ be the group generated by the c_{ij}. Every element of $\bar{G}$ is

$$c\, U_1^{x_1} U_2^{x_2} U_3^{x_3} U_4^{x_4} U_5^{x_5},$$

where c is in $\bar{C}$ and x_i is an element in $GF(p)$. The set $(x_1, x_2, x_3, x_4, x_5)$ may be taken to be a point in a projective four-space X over $\mathrm{GF}(p)$. Then every element of $\bar{G}$ not in $\bar{C}$ determines a point in X; every point in X represents a cyclic subgroup of $\bar{G}/\bar{C}$ and also an abelian subgroup of order p^{11} of $\bar{G}$. The commutator of two elements

$$c\, U_1^{x_1} \ldots U_5^{x_5} \qquad \text{and} \qquad c' U_1^{y_1} \ldots U_5^{y_5}$$

does not depend on c or c'; it is a product of the c_{ij}'s; it can be represented by a point in a projective nine-space S, over $\mathrm{GF}(p)$, which has for co-ordinates the Plücker line-co-ordinates of the line on the two points x and y in X. The points of S which represent commutators belong to V, which is a V_8^5 corresponding to the grassmannian of lines in X. Every point P of S not on V determines a three-space R in X, and the lines of R determine a five-space Σ in S, a Σ intersects V in a V_4^2; there is only one Σ in S, determined by an R in X, which contains P. A line in S which lies in a Σ is called a Σ-line; its points not on V all determine the same R in X. A line in S not a Σ-line and not intersecting V determines a unique point M on V such that the plane on M and the line is tangent to V at M. The space tangent to V at M, that is, the space consisting of all points P in S such that PM is a ruling of V or such that the five-space Σ which contains P contains M and PM meets V only at M, is six-dimensional; any plane in such a tangent space is called a τ-plane.

The four conditions on commutators which reduces $\bar{G}$ of order p^{15} to G of order p^{11} set four independent elements of $\bar{C}$ equal to identity, and hence set equal to identity all the elements in the group generated by the four. These four elements of $\bar{C}$ are represented by four independent points of S, and the points determine a three-space S_3 in S. The group G is determined by the

relation of S_3 to V. When 5 is satisfied, S_3 has no point on V; when 5 and 6 are satisfied, S_3 contains no Σ-line; when 5 and 6′ are satisfied, S_3 contains one Σ-line. These groups exist, and they are the only ones when 5 is satisfied.

The condition that S_3 intersect V leads to a fifth-degree congruence, $f(x) \equiv 0$, mod. p. If $f(x)$ has no linear factor in GF(p), condition 5 is satisfied; conditions 6 and 6′ correspond respectively to $f(x)$ irreducible and $f(x)$ the product of an irreducible quadratic and an irreducible cubic. In the latter case the quadratic is connected with the Σ-line and the cubic with a τ-plane, both unique.

3. The characteristic subgroups. A point P in S not on V determines a three-space R in X, and R determines in $\bar{G}$ a subgroup $\bar{H}$ of index p. This subgroup is the direct product of a metabelian group of order p^{10} generated by four elements and an abelian group of order p^4. When $\bar{G}$ is reduced to G by setting equal to identity the elements of $\bar{C}$ which correspond to points of S_3 $\bar{H}$ becomes a group H of order p^{10} and its commutator subgroup remains of order p^8 if P is not in the five-space Σ determined by a point of S_3; the commutator subgroup of H will be of order p^5 if P is in the Σ determined by a point of S_3 not on a Σ-line of S_3; this commutator subgroup will be of order p^4 if P is in the Σ determined by a point of the Σ-line in S_3.

When 6′ is satisfied, the unique Σ-line in S_3 means that there is one and only one three-space R in X whose Σ in S intersects S_3 in a line. Hence G contains one subgroup only of order p^{10} with commutator subgroup of order p^4; the subgroup is therefore characteristic.

The τ-plane π, which is in S_3 when 6′ is satisfied, contains no Σ-line. Hence every point of π determines a subgroup of order p^{10} in G, whose commutator subgroup has order p^5, except for the point where π intersects the Σ-line. The τ-plane is in the space tangent to V at a point M, and from this follows that M is in the five-space Σ determined by each point of π. M is the image on V of a line m in X, and m lies in every R determined by a point of π. m determines in G a subgroup H_m of order p^8 which is non-abelian since M is not in S_3. The group H_m is characterized by the fact that it is the only group of order p^8 that is contained in every one of the $1 + p + p^2$ subgroups of order p^{10} determined by the points of π. H_m is therefore characteristic in G; it is in the characteristic subgroup of order p^{10} determined by the Σ-line.

These characteristic subgroups of orders p^8 and p^{10} together with conditions $1, \ldots, 5$ are enough to determine G, and they exist for every odd p. The subgroups are seen to be characteristic because they are uniquely defined. G has other characteristic subgroups which in number depend on p, but only on the size of p. The points of the τ-plane determine subgroups of order p^{10} of G. The vertices of the frame of reference in S_3 are completely determined if S_3 is in canonical form. (**1**, p. 699). The group of isomorphisms of G determines a group of collineations of X and this group leaves invariant every point of the τ-plane and it leaves invariant two points of the Σ-line, viz., its intersection with π and the conjugate of that point with respect to the quadratic

intersection of V and the five-space Σ which contains the Σ-line. Thus every subgroup of order p^{10} determined by the points of π is characteristic.

It may be verified readily that the group of collineations of X induced by the group of isomorphisms of G is of order 2, and that the isomorphisms of G which induce the identity collineation constitute a group of order $(p-1)p^{30}$, the $p-1$ coming from replacing each U_t by its kth power, $k = 1, 2, \ldots, p-1$ and the p^{30} from replacing U_t by $c_t U_t$ where the c_t's are arbitrary, independent elements of C. The order of the group of isomorphisms is therefore $2(p-1)p^{30}$.

When 6 is satisfied S_3 contains no special line and no special plane. The relation of S_3 to V determines in S_3 p^2+1 "rational" cubic curves, one and only one through each point **(1**, pp. 704, 715–716**)**. The group of collineations of X which transforms S_3 into itself is induced by the Galois group Γ of $GF(p^5)$ relative to $GF(p)$ and hence is of order 5. Γ transforms the cubics in sets of five and hence will leave invariant a number congruent to p^2+1, mod 5; and if a cubic is invariant it will contain $p+1$, mod 5, invariant points.

The relation of S_3 to V, by which a point P determines a three-space R in X, serves to determine for any point A in R a quadric surface in S_3 which passes through P; by the same relation a point A in X but not in R determines a quadric in S_3 which does not pass through P. Thus the points of X determine in S_3 a four-parameter set W of quadrics. In X there is a locus J, of dimension three and order four, whose points determine in S_3 cones with one vertex; every point of S_3 is the vertex of one and only one such cone of the set W. Thus each point of S_3 determines a point in X, as well as the three-space R. Each point of J, in X, determines a plane σ in X which is the double tangent plane of J at the point, and intersects J in a conic C whose points determine the cones with vertices on one of the p^2+1 cubics; each of these $p+1$ cones contains the cubic curve. Two of the planes σ intersect in a point which is not on J, and this point determines a non-degenerate ruled quadric which bears the two cubics corresponding to the two planes. Moreover, every point of a plane σ not on C is on another σ.

When $p = 5t+1$, then both p^2+1 and $p+1$ are congruent to 2, nod 5, and hence S_3 contains four points fixed under Γ. These four points determine four fixed points on J, the points which determine cones with vertices at the fixed points of S_3. Moreover, X contains a fifth fixed point, the point not on J which determines the non-degenerate ruled quadric containing the two fixed cubics. These five fixed points in X determine five abelian subgroups of order p^7 in G, and these subgroups are characteristic. They are contained in sets of two, three, four in subgroups of order p^8, p^9, and p^{10}, respectively, necessarily characteristic also.

This use of the five fixed points indiscriminately does not make full use of the geometry. Four of the fixed points in X are on J and one is not. Let the fixed point not on J be A_2. A_2 is on σ_1 and σ_2, the double tangent planes of J determined by the fixed cubics K_1 and K_2 in S_3. In σ_1 and σ_2 are conics C_1 and C_2, intersections of the planes with J. C_1 and C_2 are fixed under Γ and so are

the polars l_1 and l_2 of A_2 with respect to C_1 and C_2. The other four fixed points, necessarily on J, are A_3 and A_5 on l_1 and C_1, and A_1 and A_4 on l_2 and C_2.

Distinctions can be made among the ten characteristic subgroups of order p^8 of G. We recall that each of the $1 + p + p^2 + p^3$ points of S_3 determines a unique three-space R of X and a subgroup of order p^{10} of G whose commutator subgroup has order p^8. A line in S_3 determines a line in X and also a set of $p + 1$ three-spaces in X, and a set of $p + 1$ subgroups of order p^{10} in G. Of the ten lines in X on pairs of the five points fixed under Γ, six are imaged on points of V at which the spaces tangent to V cut S_3 in lines; the remaining four do not have this property. Thus each of six of the characteristic subgroups of order p^8 is in $p + 1$ subgroups of order p^{10} whose commutator subgroups are of order p^8; each of the other four characteristic subgroups of order p^8 is in only one such group of order p^{10}.

Similar distinctions can be made among characteristic subgroups of orders p^9 and p^{10}; we will let one further example suffice. Every one of the characteristic subgroups of order p^{10} contains six of the characteristic subgroups of order p^8. The one given by $A_1 A_3 A_4 A_5$ contains the four subgroups of order p^8 described at the end of the last paragraph; each of the other four characteristic subgroups of order p^{10} contains only two of these.

When $p = 5t - 1$, S_3 contains two cubics fixed under Γ, but contains no fixed points. The cubics determine the planes σ_1 and σ_2 in X, and the intersection of σ_1 and σ_2 is a fixed point A_2. A_2 determines a characteristic subgroup of order p^7 which is abelian and the only characteristic subgroup of its order. σ_1 and σ_2 contain conics C_1 and C_2, on J, and the polars l_1 and l_2 of A_2 with respect to them. l_1 and l_2 determine characteristic subgroups of order p^8 of G. A_2 with l_1 and l_2 separately determines characteristic subgroups of order p^9, and l_1 and l_2 determine a characteristic subgroup of order p^{10}.

When $p = 5t \pm 2$, S_3 contains no fixed cubic and no fixed point. However, $1 + p + p^2 + p^3 + p^4 \equiv 1$, mod 5, and hence X contains one fixed point and one fixed three-space. Thus G contains one characteristic subgroup of each of orders p^7 and p^{10}; G contains no characteristic subgroup of order p^8 or p^9.

Thus a group of order p^{11} satisfying 1, . . . , 6 has characteristic subgroups in numbers and orders as follows:

	p^7	p^8	p^9	p^{10}
$p = 5t + 1$	5	10	10	5
$p = 5t - 1$	1	2	2	1
$p = 5t \pm 2$	1	0	0	1

4. Concluding remarks. The study of finite groups conformal with the abelian groups of orders p^n and type 1, 1, . . . , 1 immediately singles out the prime 2, since no such non-abelian group exists for $p = 2$. The maximum value of the class of a group whose elements are all of order p depends on the size of p, and so differentiates among primes; when the groups are restricted to be metabelian, that is, of class 2, this last distinction is lost. When the metabelian

groups are ordered according to the number of generators, there is no occasion to distinguish one odd prime from another until this group of order p^{11} is reached. Because of a duality in the geometry, the determination in the paper cited of all the groups of order p^{α}, $\alpha \geqslant 11$, brings with it a determination of all of those for $\alpha \leqslant 9$. There remain to be examined the groups of order p^{10}. It is certain that many of the groups of order p^{10} will require different treatment for different primes.

Reference

1. H. R. Brahana, *Metabelian p-groups with five generators and orders p^{12} and p^{11}*, Illinois J. Math., **2** (1958), 641–717.

University of Illinois

A THEOREM ON PURE SUBMODULES

GEORGE KOLETTIS, JR.

1. Introduction. In **(1)** Baer studied the following problem: If a torsion-free abelian group G is a direct sum of groups of rank one, is every direct summand of G also a direct sum of groups of rank one? For groups satisfying a certain chain condition, Baer gave a solution. Kulikov, in **(3)**, supplied an affirmative answer, assuming only that G is countable. In a recent paper **(2)**, Kaplansky settles the issue by reducing the general case to the countable case where Kulikov's solution is applicable. As usual, the result extends to modules over a principal ideal ring R (commutative with unit, no divisors of zero, every ideal principal).

The object of this paper is to carry out a similar investigation for *pure* submodules, a somewhat larger class of submodules than the class of direct summands. We ask: if the torsion-free R-module M is a direct sum of modules of rank one, is every pure submodule N of M also a direct sum of modules of rank one? Unlike the situation for direct summands, here the answer depends heavily on the ring R. If R is a field, there is no problem, and if R is a discrete valuation ring (one prime up to unit factors), it is easy to see that the answer is still yes. On the other hand, for abelian groups, or generally whenever R has an infinite number of primes, the question has a negative answer.

We fill in the gap by showing that if R has exactly two primes, an affirmative answer is obtained provided N has finite rank. If N has infinite rank or if R has three or more primes, examples are given showing that N need not be a direct sum of modules of rank one. In contrast to the large number of theorems on principal ideal rings with one prime, this appears to be the first result true specifically for rings with two primes.

2. Preliminaries. Let R be a principal ideal ring and K its quotient field. The unit of R is always assumed to act as unit operator on every R-module. We recall that a submodule N of an R-module M is *pure* if $aN = N \cap aM$ for every a in R. M is *torsion-free* if for a in R, x in M, and $ax = 0$, we have either $a = 0$ or $x = 0$. In this case, the intersection of pure submodules is pure, and so every subset of M generates a unique pure submodule. The *rank* of M is the cardinal number of a maximal set of linearly independent (over R) elements of M, or equivalently, the dimension of the K-vector space $K \otimes_R M$.

Received March 4, 1959. The author wishes to thank Professor Kaplansky for suggesting this problem. This research was supported in part by the Office of Naval Research.

The torsion-free R-modules of rank one are (up to isomorphism) the submodules of the R-module K. Two such submodules M_1 and M_2 are isomorphic if and only if $M_1 = \alpha M_2$ for some α in K. In particular, M_1 is free precisely when $M_1 = \alpha R$ for some α in K. For each prime p in R, we denote by R_p the submodule of K consisting of those elements which can be written with a denominator prime to p.

Let the torsion-free module M be a direct sum $M = \Sigma M_i$, i ranging over an index set, each M_i of rank one. Let N be a pure submodule of M. We note that we can for our purpose confine ourselves to the case where none of the summands M_i is free or divisible.

Indeed, write $M = M' \oplus F$, where F is the sum of the free M_i's and M' the sum of the remaining M_i's. $N/(M' \cap N) \cong (M' + N)/M'$ is a submodule of M/M', which is free. Thus $M' \cap N$ is a direct summand of N whose complementary summand is free. It follows that N is a direct sum of modules of rank one whenever $M' \cap N$, a pure submodule of M', is a direct sum of modules of rank one.

Next, write $M = D \oplus M''$, where D is the sum of all the divisible M_i's and M'' the sum of the remaining M_i's. The purity of N and the divisibility of D combine to yield the divisibility of $N \cap D$. Thus $N \cap D$ is a direct summand of N. The complementary summand $N/(N \cap D) \cong (N + D)/D$ is a submodule of $M/D \cong M''$. For any a in R, by the divisibility of D, we have $D = aD \subset aM$. The modular law then gives $aM \cap (N + D) = (aM \cap N) + D$, which, since N is pure, is just $aN + D$. Modulo D this becomes $(aM/D) \cap ((N + D)/D) = a(N + D)/D$, which is exactly the assertion that $(N + D)/D$ is pure in M/D. So N is a direct sum of modules of rank one whenever $N/(N \cap D)$, which can be regarded as a pure submodule of M'', is a direct sum of modules of rank one.

In conclusion, we remark that if R has just one prime, every rank one module is either free or divisible, and the above reductions are all that are needed to show that N is a direct sum of modules of rank one.

3. R with two primes. Throughout this section, we assume that R has exactly two primes (up to unit factors). Denote them by p and q. The quotient field K is the set of all fractions $a/(p^m q^n)$, a in R, $m, n \geqq 0$. The submodules of K fall into four classes according as they do or do not contain unbounded powers of p, and of q, in the denominators of their elements (when these are written in "lowest terms"). Using this classification it is easily seen that every submodule of K is isomorphic to one of R, R_p, R_q, or K. Thus these are the modules of rank one.

Now for the theorem:

THEOREM. *Let the torsion-free module M be a direct sum of modules of rank one. Then every pure submodule of finite rank is also a direct sum of modules of rank one.*

Proof. We may suppose that M has finite rank and that each rank one summand is either a copy of R_p or of R_q. Write $M = P \oplus Q$ where P is a direct sum of copies of R_p, and Q of copies of R_q. Choose elements $u_1, \ldots, u_t$ in M so that $P = R_p u_1 \oplus \ldots \oplus R_p u_s$ and $Q = R_q u_{s+1} \oplus \ldots \oplus R_q u_t$ where $0 \leqq s \leqq t$.

Assume that every pure submodule of rank $n-1$ ($n \geqq 2$) is a direct sum of modules of rank one, and let N be a pure submodule of rank n. For every k with $1 \leqq k \leqq t$, let N_k be the intersection of N with the direct sum of all the rank one summands except the kth one. Then N_k is a pure submodule of M whose rank is $n-1$ or n depending on whether or not there is an element of N having a non-zero kth component. It will be sufficient to show that at least one of the N_k of rank $n-1$ is a direct summand of N, for such an N_k is a direct sum of modules of rank one whose complementary summand is of rank one.

There is no loss in generality in assuming that $Q \neq 0$ and that $N_t \neq N$. We consider two cases:

Case I. $N \cap Q = 0$.

Since $N_t \neq N$, N_t is of rank $n-1$. We will prove that N_t is a direct summand of N by showing that N/N_t is free. To do this, we need only show that when the elements of N are expressed in terms of the u_i's there is an upper bound to the powers of p that can occur as denominators in the coefficients of u_t.

Let $x_1, \ldots, x_n$ be a maximal independent subset of N. If $x \neq 0$ is in N, some non-zero multiple of x, say rx, lies in the module generated by the x_j's. If $rx = r_1x_1 + \ldots + r_nx_n$, we can clearly suppose that not every one of $r, r_1, \ldots, r_n$ is a multiple of p in R.

Assume that N/N_t is not free, and let m be a given positive integer. Then we can choose the element x so that, in the expressions for x and the x_j's in terms of the u_i's, the coefficient of u_t for x has a power of p in its denominator so large in comparison to those for the x_j's so as to require r to be a multiple of p^m in R.

Using primes to denote images in $M/Q \cong P$, we observe that since $N \cap Q = 0$, the elements $x_1', \ldots, x_n'$ are independent. Say $x' = c_1u_1' + \ldots + c_su_s'$ and $x_j' = c_{j1}u_1' + \ldots + c_{js}u_s'$ where all the c's are in R_p. The $n \times s$ matrix (c_{ji}) thus obtained has all its rows independent. Say the first n columns are also independent, and let $D \neq 0$ be the determinant of the $n \times n$ submatrix (c_{ji}), $1 \leqq i, j \leqq n$.

We have the following system of s equations:

$$rc_i = r_1c_{1i} + \ldots + r_nc_{ni}.$$

From the first n of these, and the fact that r is in p^mR, we see that r_jD is in p^mR_p for each j. Hence D is in p^mR_p.

Thus the assumption that N/N_t is not free requires D to be in $\cap_m p^m R_p = 0$, a contradiction.

Case II. $N \cap Q \neq 0$.

Let $x \neq 0$ be an element of $N \cap Q$, say $x = b_{s+1}u_{s+1} + \ldots + b_t u_t$, each b_i in R_q. The result of dividing x by the largest power of q common to all of the b_i's will again be an element of N (N is pure), and so we may as well assume that at least one of the b_i's, say b_k, is not in qR_q.

The submodule $R_q x$ is contained in N since N is pure. We have $N = N_k \oplus R_q x$. For since $b_k \neq 0$, $N \cap R_q x = 0$. On the other hand suppose w in N has $b^*_k u_k$ as its kth component where b^*_k is in R_q. Since b_k is not in qR_q, b_k^{-1} is in R_q. Hence $w - b_k^* b_k^{-1} x$ is in N_k. This shows that $N = N_k + R_q x$.

4. Examples. Let R be an arbitrary principal ideal ring and M a torsion-free R-module. One readily verifies that for every prime p in R,

$$\bigcap_{j=1}^{\infty} p^j M$$

is a *pure* submodule of M. Since it is clear that a torsion-free module of rank one has no proper *pure* submodules, we see that if M is a direct sum ΣM_t of modules of rank one, the submodule

$$\bigcap_{j=1}^{\infty} p^j M$$

is the sum of those rank one summands M_t for which $pM_t = M_t$ and is therefore a direct summand of M. This gives a necessary condition for a module to be a direct sum of modules of rank one.

Using this condition, we give two examples. The first example shows that in the theorem the hypothesis of finite rank is indispensable. The second example shows that the theorem cannot survive the presence of three primes.

Example 1. We assume again that R has just two primes, p and q. Let $M = P \oplus Q$, where $P = R_p u_0$ is a copy of R_p and Q is a direct sum

$$\sum_{t=1}^{\infty} R_q u_t$$

of an infinite number of copies of R_q. Let N be the pure submodule generated by all the elements $(1/q^t)u_0 - u_t$. We will show that N is not a direct sum of modules of rank one.

First, we note that $N \cap P = 0$. Indeed, an element of P will lie in N only if some non-zero multiple of it lies in the module generated by the elements $(1/q^t)u_0 - u_t$. Clearly, for a_t in R, a sum

$$\sum_{t=1}^{m} a_t\left(\frac{1}{q^t} u_0 - u_t\right)$$

can only be in P if each $a_t = 0$.

Next, we note that since every element $(1/q^t)u_0$ lies in $N + Q$, we have $M = N + Q$.

Since N is pure,

$$\bigcap_{j=1}^{\infty} p^j N$$

is $N \cap Q$. Now

$$N/(N \cap Q) \cong (N + Q)/Q \cong R_p = qR_p.$$

Any submodule L of M for which $qL = L$ must be contained in P. Thus a complementary summand for $N \cap Q$ in N must be contained in $N \cap P = 0$. Since it is clear that $N \cap Q \neq N$, we conclude that $N \cap Q$ is not a direct summand of N and that N is not a direct sum of modules of rank one.

Example 2. Let R have at least three non-associated primes. Say p, q, and r are three of them. Let $M = R_p u_1 \oplus R_q u_2 \oplus R_r u_3$ be the direct sum of a copy each of R_p, R_q, and R_r. Let N be the pure submodule generated by $u_1 - u_2$ and $u_2 - u_3$. It is immediate that $N \cap R_p u_1 = 0$.

N contains all the elements $(1/p^k)(u_2 - u_3)$, $(1/q^m)(u_1 - u_3)$, and $(1/r^n)(u_1 - u_2)$. If a and b are elements of R for which $aq^m + br^n = 1$,

$$\frac{b}{q^m}(u_1 - u_3) + \frac{a}{r^n}(u_1 - u_2) - \frac{1}{q^m r^n} u_1$$

lies in $R_q u_2 + R_r u_3$. This shows that all elements of the form $(1/q^m r^n)u_1$ lie in $N + R_q u_2 + R_r u_3$. It follows that $M = N + R_q u_2 + R_r u_3$.

As in the first example,

$$\bigcap_{j=1}^{\infty} p^j N = N \cap (R_q u_2 + R_r u_3)$$

is not a direct summand of N. For

$$N/(N \cap (R_q u_2 + R_r u_3)) \cong R_p = qrR_p,$$

and a submodule L of N for which $qrL = L$ must be contained in $R_p u_1 \cap N = 0$.

References

1. R. Baer, *Abelian groups without elements of finite order*, Duke Math. J., *3* (1937), 68–122.
2. I. Kaplansky, *Projective modules*, Ann. Math., *68* (1958), 372–377.
3. L. Kulikov, *On direct decompositions of groups*, Ukrain. Mat. Z., *4* (1952), 230–275, 347–372 (Russian) = Amer. Math. Soc. Translations, Ser. 2, *2* (1956), 23–87.

University of Notre Dame

NODAL NON-COMMUTATIVE JORDAN ALGEBRAS

LOUIS. A. KOKORIS

1. Introduction. A finite dimensional power-associative algebra $\mathfrak{A}$ with a unity element 1 over a field $\mathfrak{F}$ is called a nodal algebra by Schafer **(7)** if every element of $\mathfrak{A}$ has the form $\alpha 1 + z$ where α is in $\mathfrak{F}$, z is nilpotent, and if $\mathfrak{A}$ does not have the form $\mathfrak{A} = \mathfrak{F}1 + \mathfrak{N}$ with $\mathfrak{N}$ a nil subalgebra of $\mathfrak{A}$. An algebra $\mathfrak{A}$ is called a non-commutative Jordan algebra if $\mathfrak{A}$ is flexible and $\mathfrak{A}^+$ is a Jordan algebra. Some examples of nodal non-commutative Jordan algebras were given in **(5)** and it was proved in **(6)** that if $\mathfrak{A}$ is a simple nodal non-commutative Jordan algebra of characteristic not 2, then $\mathfrak{A}^+$ is associative. In this paper we describe all simple nodal non-commutative Jordan algebras of characteristic not 2. Any such algebra has the form $\mathfrak{A} = \mathfrak{F}1 + \mathfrak{N}$ with $\mathfrak{N}^+ = \mathfrak{F}[x_1, \ldots, x_n]$ for some n where the generators are all nilpotent of index p. The x_i can be selected so that $x_ix_j = \alpha_{ij}1 + w_{ij}$ for w_{ij} in $\mathfrak{N}$ and α_{ij} in $\mathfrak{F}$ such that, for each i, some $\alpha_{ij} \neq 0$. Moreover, the multiplication table of $\mathfrak{A}$ is given by

$$(1) \qquad f(x_1, \ldots, x_n)g(x_1, \ldots, x_n) = f \cdot g + \tfrac{1}{2} \sum_{i,j} \frac{\partial f}{\partial x_i} \cdot \frac{\partial g}{\partial x_j} \cdot [x_i, x_j]$$

where the dot product $a \cdot b = \frac{1}{2}(ab + ba)$ is the product of $\mathfrak{A}^+$ and $[x_i, x_j] = x_ix_j - x_jx_i$.

The author would like to express his great indebtedness to R. D. Schafer for finding errors in the original manuscript and for showing how they could be corrected.

2. Properties of $\mathfrak{A}^+$. If $\mathfrak{D}$ is the derivation algebra of an algebra $\mathfrak{B}$, then Albert in **(1)** calls $\mathfrak{B}$ $\mathfrak{D}$-simple if there exists no ideal $\mathfrak{M}$, other than $\mathfrak{B}$ or 0, such that mD is in $\mathfrak{M}$ for every m in $\mathfrak{M}$ and D in $\mathfrak{D}$. We use a result of Harper **(2)** which for our purposes may be stated as follows.

THEOREM 1. (*Harper*) *Let $\mathfrak{B}$ be a commutative associative algebra with a unity quantity 1 over a field $\mathfrak{F}$ and let $\mathfrak{B}$ have the form $\mathfrak{B} = \mathfrak{F}1 + \mathfrak{N}$ with $\mathfrak{N}$ the radical of $\mathfrak{B}$. Also let $\mathfrak{B}$ be $\mathfrak{D}$-simple where $\mathfrak{D}$ is any set of derivations on $\mathfrak{B}$. Then $\mathfrak{N} = \mathfrak{F}[x_1, \ldots, x_n]$ for some n where the generators x_i have index p, p the characteristic of $\mathfrak{F}$.*

We remark that it is known that a $\mathfrak{D}$-simple algebra cannot have characteristic zero and Schafer has shown in **(7)** that a nodal non-commutative

Received February 17, 1958; in revised form August 31, 1959. Presented to the American Mathematical Society with the title *Nodal flexible associative-admissible algebras* on November 29, 1957.

Jordan algebra cannot have characteristic zero. He also uses a theorem of Jacobson **(4)** to prove that $\mathfrak{N}^+$ is a subalgebra of $\mathfrak{A}^+$ for any nodal non-commutative Jordan algebra.

THEOREM 2. *Let $\mathfrak{A}$ be a simple nodal non-commutative Jordan algebra over a field $\mathfrak{F}$ whose characteristic is not 2. Let $\mathfrak{D}$ be the derivation algebra of $\mathfrak{A}$. Then $\mathfrak{A}^+$ is $\mathfrak{D}$-simple.*

Suppose $\mathfrak{A}^+$ is not $\mathfrak{D}$-simple. Then there is an ideal $\mathfrak{B}$ of $\mathfrak{A}^+$ such that $\mathfrak{B}\mathfrak{D} \subseteq \mathfrak{B}$. We shall show that $\mathfrak{B}$ is then an ideal of $\mathfrak{A}$, contradicting the fact that $\mathfrak{A}$ is simple. The mapping $bD = [b, c]$ where c is any element of $\mathfrak{A}$ and $[b, c] = bc - cb$ is a derivation of $\mathfrak{A}^+$. This is so because $(a \cdot b)D = aD \cdot b + a \cdot bD$ if and only if $[a \cdot b, c] = [a, c] \cdot b + a \cdot [b, c]$ and the last identity follows from $(ab)c + (cb)a = a(bc) + c(ba)$, the linearized form of the flexible law $(ab)a = a(ba)$. Now let b be in $\mathfrak{B}$ and a in $\mathfrak{A}$. Since $\mathfrak{B}$ is a $\mathfrak{D}$-ideal of $\mathfrak{A}^+$, $bD = [b, a]$ is in $\mathfrak{B}$. Also, since $\mathfrak{B}$ is an ideal of $\mathfrak{A}^+$, $a \cdot b$ is in $\mathfrak{B}$. Then $ba - ab$ and $ab + ba$ in $\mathfrak{B}$ imply ab and ba are in $\mathfrak{B}$. That is, $\mathfrak{B}$ is an ideal of $\mathfrak{A}$.

COROLLARY. *If $\mathfrak{A} = \mathfrak{F}1 + \mathfrak{N}$ is a simple nodal non-commutative Jordan algebra over a field $\mathfrak{F}$ whose characteristic is not 2, then $\mathfrak{N}^+ = \mathfrak{F}[x_1, \ldots, x_n]$ for some n, where $x_i^p = 0$, $x_i^{p-1} \neq 0$. Thus, $\mathfrak{A}$ has order p^n.*

3. The multiplication table of $\mathfrak{A}$. Assume that $\mathfrak{A}$ is simple so that, by the corollary above, $\mathfrak{A}^+ = \mathfrak{F}[1, x_1, \ldots, x_n]$ with $x_i^p = 0$. In **(3)**, Jacobson has shown that if D is any derivation on $\mathfrak{A}^+$, then

$$fD = \sum_i \frac{\partial f}{\partial x_i} \cdot a_i$$

for any f in $\mathfrak{A}^+$ and for a_i in $\mathfrak{A}^+$. The a_i of course depend on the derivation D. If g is any element of $\mathfrak{A}^+$, we have seen that the mapping $fD = [f, g]$ is a derivation of $\mathfrak{A}^+$. Hence

$$fD = [f, g] = \sum_i \frac{\partial f}{\partial x_i} \cdot a_i(g).$$

To evaluate the $a_i(g)$, we note that $x_iD = [x_i, g] = a_i(g)$ and

$$[g, x_i] = \sum_j \frac{\partial g}{\partial x_j} \cdot a_j(x_i).$$

Since $[x_i, g] = -[g, x_i]$,

$$a_i(g) = -\sum_j \frac{\partial g}{\partial x_j} \cdot a_j(x_i)$$

and since $[x_j, x_i] = a_j(x_i)$, it follows that

$$[f, g] = \sum_{i,j} \frac{\partial f}{\partial x_i} \cdot \frac{\partial g}{\partial x_j} \cdot [x_i, x_j].$$

THEOREM 3. *If $\mathfrak{A}$ is a simple algebra, then for any f, g in $\mathfrak{A}$,*

$$fg = f \cdot g + \tfrac{1}{2} \sum_{i,j} \frac{\partial f}{\partial x_i} \cdot \frac{\partial g}{\partial x_j} \cdot [x_i, x_j].$$

This result follows from the above formula for $[f, g]$ and the fact that $fg = f \cdot g + \frac{1}{2}[f, g]$. The assumption that $\mathfrak{A}$ is nodal implies that at least one of the $[x_i, x_i]$ is not in $\mathfrak{N}$. This is equivalent to the statement that for some i, j, $x_i x_j$ is not in $\mathfrak{N}$.

THEOREM 4. *The generators $x_1, \ldots, x_n$ can be selected so that $x_i x_j = \alpha_{ij} 1 + w_{ij}$ with w_{ij} in $\mathfrak{N}$ and α_{ij} in $\mathfrak{F}$ such that, for each i, some $\alpha_{ij} \neq 0$.*

Let $\mathfrak{M}$ be the vector space with $x_1, \ldots, x_n$ as a basis. If we write $\alpha_{ij} = \alpha(x_i, x_j)$ then $x_j x_i = 2x_i \cdot x_j - x_i x_j = -\alpha_{ij} - w_{ij} + 2x_i \cdot x_j$ together with the fact that $x_i \cdot x_j$ is in $\mathfrak{N}$, implies that $\alpha(x_j, x_i) = -\alpha(x_i, x_j)$. Therefore $\alpha(x_i, x_j)$ is a skew-symmetric bilinear form on $\mathfrak{M}$. If the rank of the form is $2r$, there exists a basis $x_1', \ldots, x_n'$ such that we have the canonical form

$$\alpha(x_i', x_{i+r}') = 1 = -\alpha(x_{i+r}', x_i')$$

for $i \leqslant r$, $\alpha(x_i', x_j') = 0$ for all other pairs i, j. Next take $x_i'' = x_i'$ for $i \leqslant 2r$ and $x_i'' = x_i' + x_1'$ for $i > 2r$. Then, if $i \leqslant r$, $\alpha(x_i'', x_{i+r}'') = \alpha(x_i', x_{i+r}') = 1$; if $r < i \leqslant 2r$,

$$\alpha(x_i'', x_{i-r}'') = \alpha(x_i', x_{i-r}') = -\alpha(x_{i-r}', x_{(i-r)+r}') = -1;$$

and if $i > 2r$, $\alpha(x_i'', x_{r+1}'') = \alpha(x_i' + x_1', x_{r+1}') = \alpha(x_1', x_{r+1}') = 1$. The basis $x_1'', \ldots, x_n''$ of $\mathfrak{M}$ has the properties stated in Theorem 4.

4. Construction of algebras. Let $\mathfrak{F}$ be any field of characteristic $p \neq 2$. Define $\mathfrak{A}^+$ by $\mathfrak{A}^+ = \mathfrak{F}1 + \mathfrak{N}^+$ where $\mathfrak{N}^+ = \mathfrak{F}[x_1, \ldots, x_n]$ with $x_1, \ldots, x_n$ nilpotent generators of index p. That is, $\mathfrak{A}^+$ consists of elements $\alpha 1 + z$ where α is in F, 1 is the unity quantity of $\mathfrak{A}^+$, and z is a polynomial in $x_1, \ldots, x_n$. Define the algebra $\mathfrak{A} = \mathfrak{F}1 + \mathfrak{N}$ to be the same vector space as $\mathfrak{A}^+$ and to have a product defined by $x_i x_j = \alpha_{ij} 1 + w_{ij}$ for any $\alpha_{ij} = -\alpha_{ji}$ in $\mathfrak{F}$ and and $w_{ij} = 2x_i \cdot x_j - w_{ji}$ in $\mathfrak{N}$, $i < j$. Further define

$$fg = f \cdot g + \tfrac{1}{2} \sum_{i,j} \frac{\partial f}{\partial x_i} \cdot \frac{\partial g}{\partial x_j} \cdot [x_i, x_j]$$

for f, g any elements in $\mathfrak{A}$.

THEOREM 5. *If at least one $\alpha_{ij} \neq 0$, the algebra $\mathfrak{A}$ described above is a nodal non-commutative Jordan algebra.*

Linearization of the flexible law $(fg)f = f(gf)$ yields the identity $(fg)h + (hg)f = f(gh) + h(gf)$. Add $(gf)h + (gh)f$ to both sides of the equality to obtain

$$(f \cdot g)h + (g \cdot h)f = (gf) \cdot h + (gh) \cdot f. \tag{2}$$

Since $\mathfrak{A}$ has characteristic $\neq 2$, flexibility is equivalent to identity (2). The expression

$$gf\cdot h + gh\cdot f - (g\cdot h)f - (f\cdot g)h$$

$$= f\cdot g\cdot h + \tfrac{1}{2}\sum_{i,j}\frac{\partial g}{\partial x_i}\cdot\frac{\partial f}{\partial x_j}\cdot[x_i, x_j]\cdot h + f\cdot g\cdot h$$

$$+ \tfrac{1}{2}\sum_{i,j}\frac{\partial g}{\partial x_i}\cdot\frac{\partial h}{\partial x_j}\cdot[x_i, x_j]\cdot f - f\cdot g\cdot h$$

$$- \tfrac{1}{2}\sum_{i,j}\frac{\partial (g\cdot h)}{\partial x_i}\cdot\frac{\partial f}{\partial x_j}\cdot[x_i, x_j] - f\cdot g\cdot h$$

$$- \tfrac{1}{2}\sum_{i,j}\frac{\partial (f\cdot g)}{\partial x_i}\cdot\frac{\partial h}{\partial x_j}\cdot[x_i, x_j].$$

Using

$$\frac{\partial (a\cdot b)}{\partial x} = \frac{\partial a}{\partial x}\cdot b + a\cdot\frac{\partial b}{\partial x},$$

the above expression becomes

$$\tfrac{1}{2}\sum_{i,j}[x_i, x_j]\cdot\left(\frac{\partial g}{\partial x_i}\cdot\frac{\partial f}{\partial x_j}\cdot h + \frac{\partial g}{\partial x_i}\cdot\frac{\partial h}{\partial x_j}\cdot f\right.$$

$$\left. - \frac{\partial g}{\partial x_i}\cdot\frac{\partial f}{\partial x_j}\cdot h - \frac{\partial h}{\partial x_i}\cdot\frac{\partial f}{\partial x_j}\cdot g - \frac{\partial f}{\partial x_i}\cdot\frac{\partial h}{\partial x_j}\cdot g - \frac{\partial g}{\partial x_i}\cdot\frac{\partial h}{\partial x_j}\cdot f\right)$$

$$= \tfrac{1}{2}\sum_{i,j}[x_i, x_j]\cdot\left(-\frac{\partial h}{\partial x_i}\cdot\frac{\partial f}{\partial x_j} - \frac{\partial f}{\partial x_i}\cdot\frac{\partial h}{\partial x_j}\right)\cdot g$$

$$= f\cdot g\cdot h - (hf)\cdot g + f\cdot g\cdot h - (fh)\cdot g = 0$$

as desried. The algebra is nodal since at least one α_{ij} is not zero.

The proof of Theorem 4 depends only on $\mathfrak{A}$ having the form as described at the beginning of this section and it is not necessary for $\mathfrak{A}$ to be simple in order to obtain the result of Theorem 4. Thus we may assume that the generators $x_1, \ldots, x_n$ have the properties of Theorem 4 and that we have the associated bilinear form of rank $2r$.

THEOREM 6. *If $n = 2r$, then $\mathfrak{A}$ is simple.*

Suppose $\mathfrak{B}$ is a proper ideal of $\mathfrak{A}$. Then there exists a polynomial $f = f(x_1, \ldots, x_n)$ in $\mathfrak{B}$ with least possible degree t in $x_1, \ldots, x_n$. Since $n = 2r$, $\alpha_{ij} = 0$ except for the following: $\alpha_{i,r+i} = 1$ for $i \leqslant r$; and $\alpha_{i,i-r} = -1$ for $r < i \leqslant 2r$. Then for each i there exists a k such that $\alpha_{ki} \neq 0$ but $\alpha_{kj} = 0$ for all $j \neq i$. Then for this i,

$$x_k f = \sum_j \alpha_{kj}\frac{\partial f}{\partial x_j} + \text{terms of degree} \geqslant t = \alpha_{ki}\frac{\partial f}{\partial x_i} + \text{terms of degree} \geqslant t.$$

Therefore, if any monomial of f of degree t has a power x_i as a factor, $x_k f$ is a polynomial of degree $t-1$. The fact that f is in $\mathfrak{B}$ implies that $x_k f$ is in $\mathfrak{B}$ and this contradicts the assumption that f has minimal degree t.

If $n > 2r$, $\mathfrak{A}$ is not necessarily simple. For example, consider $x_1 - x_{2r+1}$ which has the property that $(x_1 - x_{2r+1})\mathfrak{A} \subseteq \mathfrak{N}$. Then $\mathfrak{B} = (x_1 - x_{2r+1})\cdot\mathfrak{A}$ is an ideal of $\mathfrak{A}$ if

$$[(x_1 - x_{2r+1})\cdot g]f = (x_1 - x_{2r+1})\cdot g\cdot f + \tfrac{1}{2}\sum_{i,j}\frac{\partial[(x_1 - x_{2r+1})\cdot g]}{\partial x_i}\cdot\frac{\partial f}{\partial x_j}\cdot[x_i, x_j]$$

$$= (x_1 - x_{2r+1})\cdot g\cdot f + \tfrac{1}{2}\sum_{j}\frac{\partial f}{\partial x_j}\cdot g\cdot[x_1 - x_{2r+1}, x_j]$$

$$+ \tfrac{1}{2}\sum_{i,j}\frac{\partial g}{\partial x_i}\cdot\frac{\partial f}{\partial x_j}\cdot[x_i, x_j]\cdot(x_1 - x_{2r+1})$$

is in $\mathfrak{B}$ for every g and f in $\mathfrak{A}$. This will be so if $[x_1 - x_{2r+1}, x_j]$ is in $\mathfrak{B}$ for every j. This can be accomplished by setting $x_1x_j = x_jx_1 = x_1\cdot x_j$ and $x_{2r+1}x_j = x_jx_{2r+1} = x_{2r+1}\cdot x_j$. Then $[x_1 - x_{2r+1}, x_j] = 0$ is certainly in $\mathfrak{B}$ for every j.

It seems clear that whether or not $\mathfrak{A}$ is simple with $n > 2r$ depends on the nature of the nilpotent elements w_{ij}.

References

1. A. A. Albert, *On commutative power-associative algebras of degree two*, Trans. Amer. Math. Soc., *74* (1953), 323–343.
2. L. R. Harper, *Some properties of partially stable algebras*, University of Chicago Ph.D. dissertation.
3. N. Jacobson, *Classes of restricted Lie algebras of characteristic p. II*, Duke Math. J., *10* (1943), 107–121.
4. ——— *A theorem on the structure of Jordan algebras*, Proc. Nat. Acad. Sci. U.S.A., *42* (1956), 140–147.
5. L. A. Kokoris, *Some nodal noncommutative Jordan algebras*, Proc. Amer. Math. Soc., *9* (1958), 164–166.
6. ——— *Simple nodal noncommutative Jordan algebras*. Proc. Amer. Math. Soc., *9* (1958), 652–654.
7. R. D. Schafer, *On noncommutative Jordan algebras*, Proc. Amer. Math. Soc., *9* (1958), 110–117.

Illinois Institute of Technology

GENERALIZED LIE ELEMENTS

RIMHAK REE

Introduction. Let $\lambda(ij)$, $i, j = 1, 2, \ldots, m$, be m^2 elements in a field K of characteristic zero such that $\lambda(ij)\lambda(ji) = 1$ for all i and j, and $x_1, x_2, \ldots, x_m$ non-commutative associative indeterminates over K. Define the elements $[x_{i_1}x_{i_2}\ldots x_{i_n}]$ inductively by $[x_i] = x_i$ and

$$[x_{i_1}x_{i_2}\ldots x_{i_n}] = x_{i_1}[x_{i_2}\ldots x_{i_n}] - \prod_{r=2}^{n} \lambda(i_1 i_r)[x_{i_2}\ldots x_{i_n}]x_{i_1}.$$

Any linear combination of the elements

$$[x_{i_1}x_{i_2}\ldots x_{i_n}]$$

with coefficients in K will be called a *generalized Lie elememt*. Generalized Lie elements reduce to ordinary Lie elements if $\lambda(ij) = 1$ for all i and j.

The purpose of this paper is to generalize to the generalized Lie elements the following: a theorem of Friedrichs, a theorem of Dynkin-Specht-Wever **(2)**, and the Witt formula on the dimension of the space spanned by homogeneous Lie elements of a fixed degree. The set of all generalized Lie elements will be made into an algebra which generalizes the ordinary free Lie algebra. This algebra turns out to be free in a certain sense. We shall also generalize the algebra associated with shuffles in **(2)**.[1]

1. Generalized Lie algebras. Throughout this paper K will denote a field of characteristic zero. By a *bi-character* in K of an additively written abelian semi-group M we shall mean a map $\chi: M \times M \to K$ satisfying the following:

$$\chi(\rho, \sigma + \tau) = \chi(\rho, \sigma)\chi(\rho, \tau), \chi(\rho + \sigma, \tau) = \chi(\rho, \tau)\chi(\sigma, \tau)$$

for all ρ, σ, τ in M. A bi-character χ will be called *skew-symmetric* if $\chi(\sigma, \tau)$ $\chi(\tau, \sigma) = 1$ for all σ, τ in M. An (associative or non-associative) algebra A over K is said to be *graded* by the semi-group M if A is a direct sum of subspaces A_ρ indexed by $\rho \in M$ such that $f \in A_\rho$ and $g \in A_\sigma$ imply $fg \in A_{\rho+\sigma}$.

Let L be an algebra graded by M, and let χ be a skew-symmetric bi-character of M in K. We shall call L a *generalized Lie algebra of type* χ, or simply a χ-algebra, if $f \in L_\rho$, $g \in L_\sigma$, imply

$$[f, g] + \chi(\rho, \sigma)[g, f] = 0;$$

Received March 30, 1959.

[1]The referee remarks that the algebras considered in this paper include, as a special case, the "left Lie algebras" which are used in homological algebra (cf. for example, the exposition by P. Cartier in Séminaire Bourbaki, May, 1955).

$$[f, [g, h]] - \chi(\rho, \sigma)[g, [f, h]] = [[f, g], h],$$

where $[f, g]$ denotes the product in L of f and g. In case χ is trivial, a χ-algebra is clearly an ordinary Lie algebra. Let A be an associative algebra graded by M. Define a new multiplication $[a, b]$ in the vector space A by

$$[a, b] = ab - \chi(\rho, \sigma)ba,$$

where $a \in A_\rho$, $b \in A_\sigma$. Then we obtain a new algebra which we shall denote by $[A]$. It can be seen easily that $[A]$ is a χ-algebra.

Let L and L' be two algebras graded by the same M. A linear map $\phi: L \to L'$ will be said to *respect grade* if $f \in L_\rho$ implies $\phi(f) \in L_\rho'$. Let L be a χ-algebra and A an associated algebra both graded by M. A grade-respecting linear map $\phi: L \to A$ will be called a *linearization* of L in A if ϕ is a homomorphism of L into $[A]$, that is, if

$$\phi([f, g]) = \phi(f)\phi(g) - \chi(\rho, \sigma)\phi(g)\phi(f)$$

for all $f \in L_\rho$, $g \in L_\sigma$. The tensor algebra T over the vector space L is graded by M if T_ρ is defined to be the subspace spanned by elements of the form $f_1 \otimes f_2 \otimes \ldots \otimes f_n$, where $f_i \in L_{\rho_i}$ and $\rho_1 + \rho_2 + \ldots + \rho_n = \rho$. Let J be the two-sided ideal of T generated by homogeneous elements of the form $f \otimes g - \chi(\rho, \sigma)g \otimes f - [f, g]$, where $f \in L_\rho$, $g \in L_\sigma$. Then the algebra $U = T/J$ is also graded by M, and the inclusion map $L \to T$ induces a linearization η of L in U. The algebra U will be called the *universal enveloping algebra* of L; it can be characterized by the property: for any linearization $\phi: L \to A$ of L into an associative algebra A, there exists a grade-respecting homomorphism $\xi: U \to A$ such that $\phi = \xi \circ \eta$.

2. Finitely generated free χ-algebras. From now on we shall consider χ-algebras L satisfying the following conditions (2.1) – (2.4):

(2.1) M is a free abelian group of rank m, with basis elements $\rho_1, \rho_2, \ldots, \rho_m$;

(2.2) $L_\rho = 0$ unless ρ is of the form $\rho = t_1\rho_1 + t_2\rho_2 + \ldots + t_m\rho_m$, where $t_1, t_2, \ldots, t_m$ are non-negative integers not all of which are zero;

(2.3) each L_{ρ_i} $(i = 1, 2, \ldots, m)$ is of dimension 1;

(2.4) L is generated by $L_{\rho_1}, L_{\rho_2}, \ldots, L_{\rho_m}$.

A χ-algebra L satisfying (2.1) – (2.4) above, will be called a *free χ-algebra of rank m* if any χ-algebra satisfying (2.1) – (2.4) is a (grade-respecting) homomorphic image of L. The existence of a free χ-algebra can be seen as follows: let F be the free (non-associative) algebra generated by an m-dimensional vector space E over the field K. If we choose a basis of E over K, then F can be graded in an obvious way by the free abelian group M of rank m. Let J be the two-sided ideal of F generated by homogeneous elements of the forms $fg + \chi(\rho, \sigma)gf$ and $f(gh) - \chi(\rho, \sigma)g(fh) - (fg)h$, where $f \in F_\rho$, $g \in F_\sigma$. Then $L = F/J$ is easily seen to be a free χ-algebra of rank m.

Let U be the universal enveloping algebra of the free χ-algebra L of rank m with the linearization map $\eta: L \to U$, and let A be the free associative algebra over K generated by m free generators $x_1, x_2, \ldots, x_m$. Since L is free, there exists a homomorphism $\phi: L \to [A]$ such that $\phi(f_i) = x_i$, $i = 1, 2, \ldots, m$, that is, ϕ is a linearization of L in A. Then by the definition of U, there exists a grade-respecting homomorphism $\xi: U \to A$ such that $\phi = \xi \circ \eta$. Then ξ must be an isomorphism, since A is free-associative. Thus we may regard U as a free associative algebra with free generators $x_1 = \eta(f_1), \ldots, x_m = \eta(f_m)$. The fact that $\eta(f) = 0$ implies $F = 0$ can also be proved in exactly the same way as in the case of free Lie algebras **(3, 1–9)**. Hence we may identify L as the subalgebra of $[U]$ generated by $x_1, \ldots, x_m$. It can be seen easily that L is spanned by the elements

$$[x_{i_1}x_{i_2}\ldots x_{i_n}] = [x_{i_1}[\ldots[x_{i_{n-1}}x_{i_n}]\ldots]]$$

defined in the Introduction by using $\lambda(ij) = \chi(\rho_i, \rho_j)$. Thus we may state

THEOREM 2.5. *Let K be a field of characteristic zero, $x_1, x_2, \ldots, x_m$ non-commutative associative indeterminates over K, and $\lambda(ij)$, $i, j = 1, 2, \ldots, m$, be m^2 elements in K such that $\lambda(ij)\lambda(ji) = 1$ for all i and j. Then the vector space over K spanned by the elements*

$$[x_{i_1}x_{i_2}\ldots x_{i_n}]$$

defined above forms a free χ-algebra L with respect to the multiplication

$$[[x_{i_1}\ldots x_{i_p}], [x_{j_1}\ldots x_{j_q}]] = [x_{i_1}\ldots x_{i_p}][x_{j_1}\ldots x_{j_q}] - \prod_{\mu=1}^{p}\prod_{\nu=1}^{q}\lambda(i_\mu j_\nu)[x_{j_1}\ldots x_{j_q}][x_{i_1}\ldots x_{i_p}].$$

The universal enveloping algebra of L is isomorphic to the free associative algebra with m free generators.

It should be understood in the above theorem that L is graded by M as follows: for $\rho = t_1\rho_1 + t_2\rho_2 + \ldots + t_m\rho_m$, L_ρ consists of linear combinations of elements of the form

$$[x_{i_1}x_{i_2}\ldots x_{i_n}]$$

in which, for each i, x_i appears t_i times. Also, χ is defined by $\chi(\rho_i, \rho_j) = \lambda(ij)$.

3. A generalization of a Witt formula. Let L be as in Theorem 2.5. An element in L will be called a *homogeneous element of degree n* if it is a linear combination of elements of the form

$$[x_{i_1}x_{i_2}\ldots x_{i_n}].$$

In this section we shall compute the dimension of the space spanned by all homogeneous elements of degree n, following a method given by Witt **(4)**. By the same method one may be able to compute the dimension of each L_ρ.

Let A and B be two associative algebras both graded by M, and $A \otimes B$ the tensor product of A and B regarded as vector spaces over K. Using a bi-character χ of M, define a multiplication in the vector space $A \otimes B$ by

$$(a \otimes b)(a' \otimes b') = \chi(\sigma, \rho')(aa' \otimes bb')$$

where $b \in B_\sigma$, $a' \in A_\rho'$. The algebra obtained in this way is easily seen to be associative, and will be denoted simply by $A \otimes B$. It will be used in the proof of (3.1), below, as well as in the formulation of a generalization of a theorem of Friedrichs.

Now, for the skew-symmetric bi-character χ of M, we have $\chi(\rho, \rho) = \pm 1$ for any $\rho \in M$. The subspace L_ρ of the free χ-algebra L will be called *positive* or *negative* according as $\chi(\rho, \rho) = 1$ or $\chi(\rho, \rho) = -1$. Choose a basis for each positive L_ρ and let the union of these basis elements be $P_1, P_2, P_3 \ldots$. Also, choose a basis for each negative L_ρ and let the union of these basis elements be $Q_1, Q_2, Q_3 \ldots$. Let $\eta: L \to U$ be the linearization of L into its universal enveloping algebra U. Then we have

THEOREM 3.1. *The elements*

$$\eta(P_1)^{s_1}\, \eta(P_2)^{s_2} \ldots \eta(P_k)^{s_k}\, \eta(Q_1)^{t_1}\, \eta(Q_2)^{t_2} \ldots \eta(Q_n)^{t_n}$$

form a basis of the universal enveloping algebra U of the free χ-algebra L. Here the indices run as follows: $s_1, s_2 \ldots$ are non-negative integers; each of t_i is either 0 or 1; $k, n = 0, 1, 2 \ldots$.

Proof. Since, for each i,

$$\eta([Q_i, Q_i]) = \eta(Q_i)^2 - \chi(\rho, \rho)\eta(Q_i)^2 = 2\eta(Q_i)^2,$$

it follows that $\eta(Q_i)^2$ is a linear combination of some $\eta(P_j)$'s and some $\eta(Q_k)$'s. Then by the definition of the linearization, it is clear that U is spanned by the given elements. Thus it remains to show that the given elements are linearly independent. For this purpose, let U' be a replica of U with grade-respecting isomorphism $\iota: U \to U'$, and let $\eta' = \iota \circ \eta$. Let $U \otimes U'$ be the tensor product of U and U' with respect to χ. Then $U \otimes U'$ is also graded by M in an obvious way, and the map $\bar{\eta}: L \to U \otimes U'$ defined by

$$\bar{\eta}(f) = \eta(f) \otimes 1 + 1 \otimes \eta'(f)$$

is easily seen to be a linearization of L into $U \otimes U'$. Therefore there exists a homomorphism $\xi: U \to U \otimes U'$ such that $\xi \circ \eta = \bar{\eta}$. Using ξ, one may now prove the linear independence of the given elements in exactly the same way as in the case of ordinary Lie algebras **(3**, pp. 1–8**)**. We omit the details.

Now, let the free χ-algebra L given in (2.5) be graded by M as in the remark following (2.5). Let the basis elements $\rho_1, \rho_2, \ldots, \rho_m$ be such that

$$L_{\rho_1}, \ldots, L_{\rho_p}$$

are positive while

$$L_{\rho_{p+1}}, \ldots, L_{\rho_{p+q}}$$

$(p + q = m)$ are negative. Since, for $\rho = t_1\rho_1 + \ldots + t_m\rho_m$,

$$\chi(\rho, \rho) = \prod_{i,j} \chi(\rho_i, \rho_j)^{t_i t_j} = \prod_i \chi(\rho_i, \rho_i)^{t_i^2} = (-1)^t,$$

where $t = t_{p+1} + \ldots + t_{p+q}$, it follows that

$$[x_{i_1} x_{i_2} \ldots x_{i_n}]$$

belongs to a positive L_ρ if and only if its degree with respect to $x_{p+1}, \ldots, x_{p+q}$ is even. Denote by p_n and q_n, respectively, the numbers of P_i's of degree n and the numbers of Q_i's of degree n, and consider the formal power series

$$F(x, \lambda) = \prod_{d=1}^{\infty} (1 + x^d + x^{2d} + \ldots)^{p_d}(1 + \lambda x^d)^{q_d}$$

with a parameter λ. The coefficient $c_n(\lambda)$ of x^n in $F(x)$ is a polynomial in λ with integral coefficients. By (3.1), $c_n(1)$ is equal to the dimension of the subspace of U spanned by all homogeneous elements of degree n; $c_n(1) = (p + q)^n$. On the other hand, also by (3.1), $c_n(-1) = a_n - b_n$, where a_n denotes the dimension of the subspace A_n of U spanned by all homogeneous elements which are of even degrees with respect to $x_{p+1}, \ldots, x_{p+q}$, and where b_n denotes the dimension of the subspace B_n of U spanned by all homogeneous elements which are of odd degrees with respect to $x_{p+1}, \ldots, x_{p+q}$. Since U is free associative, A_n (resp. B_n) is spanned by elements

$$x_{i_1} x_{i_2} \ldots x_{i_n}$$

of even (resp. odd) degree with respect to $x_{p+1}, \ldots, x_{p+q}$. Thus

$$a_n = C_{n,0}p^n + C_{n,2}p^{n-2}q^2 + \ldots,$$

$$b_n = C_{n,1}p^{n-1}q + C_{n,3}p^{n-3}q^3 + \ldots,$$

where $C_{n,r}$ are binomial coefficients. Hence $a_n - b_n = (p - q)^n$, and we have

$$F(x, 1) = 1 + (p + q)x + (p + q)^2x^2 + \ldots,$$

$$F(x, 1) = -1 + (p - q)x + (p - q)^2x^2 + \ldots.$$

Taking logarithms of both sides, and comparing the coefficients of x^n/n, we have, for $n = 1, 2, \ldots$,

$$\sum_{d|n} dp_d - \sum_{d|n} (-1)^{n/d} dq_d = (p + q)^n,$$

$$\sum_{d|n} dp_d - \sum_{d|n} dq_d = (p - q)^n.$$

Let $k > 0$ be an odd integer. Then, since

$$\sum_{d|2^\alpha k} (-1)^{2^\alpha k/d} dq_d = \sum_{d|2^{\alpha-1}k} dq_d - \sum_{d|k} 2^\alpha dq_{2^\alpha d},$$

$$\sum_{d|2^\alpha k} dp_d = \sum_{d|2^{\alpha-1}k} dp_d + \sum_{d|k} 2^\alpha dp_{2^\alpha d}$$

we obtain, from the above,

$$\sum_{d|k} 2^{\alpha} d(p_{2^{\alpha}d} + q_{2^{\alpha}d}) = (p+q)^{2^{\alpha}k} - (p-q)^{2^{\alpha}k}.$$

Then by the Möbius inversion formula, we have

$$p_{2^{\alpha}k} + q_{2^{\alpha}k} = \frac{1}{2^{\alpha}k} \sum_{d|k} \mu(d)((p+q)^{2^{\alpha}k/d} - (p-q)^{2^{\alpha}k/d}).$$

In case $\alpha = 0$, the above reduces (for odd k) to

$$p_k + q_k = \frac{1}{k} \sum_{d|k} \mu(d)(p+q)^{k/d}$$

Following Witt, we shall use the notations:

$$\psi(n) = \frac{1}{n} \sum_{d|n} \mu(d)(p+q)^{n/d};$$

$$\psi^*(n) = p_n + q_n.$$

Then the above can be summarized as

THEOREM 3.2. *The dimension $\psi^*(n)$ of the vector space spanned by all elements of the form*

$$[x_{i_1} x_{i_2} \ldots x_{i_n}]$$

is given, for odd k, by

$$\psi^*(k) = \psi(k);$$

$$\psi^*(2^{\alpha}k) = \psi(2^{\alpha}k) + \frac{1}{2^{\alpha}k} \sum_{d|k} \mu(d)((p+q)^{2^{\alpha-1}k/d} - (p-q)^{2^{\alpha-1}k/d}),$$

where p denotes the number of indices i such that $\lambda(ii) = \chi(\rho_i, \rho_i) = 1$ while q denotes the number of indices j such that $\lambda(jj) = -1$.

It should be remarked that the function $\psi^*(n)$ is completely determined by the values of $\lambda(ii)$, and independent of other values of $\lambda(ij)$. The Witt formula is obtained as the case $q = 0$. In case all $\lambda(ii) = -1$, we have $p = 0$, and we may deduce from the above that

$$\psi^*(n) = \begin{cases} \psi(n) & \text{for } n \equiv 0, 1, 3 \pmod 4, \\ \psi(n) + \psi(\tfrac{1}{2}n) & \text{for } n \equiv 2 \pmod 4. \end{cases}$$

4. An algebra associated with shuffles. We shall generalize the algebra defined in **(2)** to apply to generalized Lie elements. If r and s are positive integers, define a *shuffle of type* (r, s) to be a permutation σ of the numbers $1, 2, \ldots, r+s$ such that $1 \leqslant \sigma(\mu) < \sigma(\nu) \leqslant r$ or $r < \sigma(\mu) < \sigma(\nu) \leqslant r+s$ implies $\mu < \nu$. Take m^2 elements $\lambda(ij)$ in K arbitrarily, and define an algebra A over K as follows. A has the basis

$$\{1\} \cup \{a(i_1 \ldots i_n) | i_1, \ldots, i_n = 1, 2, \ldots, m;\ n = 1, 2, \ldots\}$$

with the multiplication table: 1 is a unity element;

$$a(i_1 \ldots i_r)a(i_{r+1} \ldots i_{r+s}) = \sum_{\sigma} \lambda(\sigma) a(i_{\sigma(1)} i_{\sigma(2)} \ldots i_{\sigma(r+s)}),$$

where the sum ranges over all shuffles σ of type (r, s) while $\lambda(\sigma)$ denotes the product of all $\lambda(i_{\sigma(\mu)}, i_{\sigma(\nu)})$ such that $\mu < \nu$ and $\sigma(\mu) > \sigma(\nu)$. (We set $\lambda(\sigma) = 1$ if σ is the identity permutation.)
Thus, for example,

$$a(i)a(j) = a(ij) + \lambda(ji)a(ji);$$
$$a(i)a(jk) = a(ijk) + \lambda(ji)a(jik) + \lambda(ji)\lambda(ki)a(jki).$$

THEOREM 4.1. *The algebra A is associative, and if $\lambda(ij)\lambda(ji) = 1$ for all i and j, then it satisfies the generalized commutativity:*

$$a(j_1 \dots j_s)a(i_1 \dots i_r) = \prod_{\mu=1}^{r} \prod_{\nu=1}^{s} \lambda(i_\mu j_\nu)a(i_1 \dots i_r)a(j_1 \dots j_s).$$

Proof. If

$$f = a(i_1 \dots i_r),\ g = a(i_{r+1} \dots i_{r+s}),\ h = a(i_{r+s+1} \dots i_{r+s+t}).$$

then it is readily seen that both $(fg)h$ and $f(gh)$ are of the form

$$\sum \lambda(\sigma)\, a(i_{\sigma(1)}i_{\sigma(2)} \dots i_{\sigma(r+s+t)}),$$

where σ runs over all permutations of $1, 2, \dots, r + s + t$ such that any one of the three conditions

$$1 \leqslant \sigma(\mu) < \sigma(\nu) < r,$$
$$r < \sigma(\mu) < \sigma(\nu) \leqslant r + s,$$
$$r + s < \sigma(\mu) < \sigma(\nu) < r + s + t$$

implies $\mu < \nu$, and where $\lambda(\sigma)$ denotes the product of all $\lambda(i_{\sigma(\mu)}i_{\sigma(\nu)})$ such that $\mu < \nu$ and $\sigma(\mu) > \sigma(\nu)$. Hence $(fg)h = f(gh)$. The second half of the theorem may be verified easily.

In the rest of this section, we shall assume that $\lambda(ij)\lambda(ji) = 1$ for all i and j. Making the convention that $a(i_1 \dots i_r)$ stands for 1 whenever the set $\{i_1, \dots, i_r\}$ of indices is empty, we define the bilinear operation $\vee$ in A by

$$a(i_1 \dots i_r) \vee a(j_1 \dots j_s) = a(i_1 \dots i_r\ j_1 \dots j_s).$$

We also make the convention that the multiplication in A has priority over the operation $\vee$.

Define the elements $a[i_1 i_2 \dots i_n]$ in A inductively by $a[i] = a(i)$ and

$$a[i_1 i_2 \dots i_n] = a(i_1) \vee a[i_2 \dots i_n] - \prod_{\nu=1}^{n-1} \lambda(i_n i_\nu)a(i_n) \vee a[i_1 \dots i_{n-1}].$$

For the generalizations in the next section of some theorems on Lie elements, we need the following

THEOREM 4.2. *For $n > 0$, we have*

$$\sum_{s=1}^{n} a[i_1 \dots i_s]a(i_{s+1} \dots i_n) = na(i_1 \dots i_n).$$

The above theorem may be proved in exactly the same way as in the case where all $\lambda(ij) = 1$ **(2)**, if we use the linear map $D: A \to A$ defined by $D(1) = 0$ and

$$Da(i_1 i_2 \dots i_n) = \gamma(i_1) a(i_2 \dots i_n),$$

where $\gamma(1), \dots, \gamma(m)$ are m arbitrary elements in K. We omit the proof of (4.2). Incidentally, the map D becomes an anti-derivation of A if all $\lambda(ij) = -1$.

THEOREM 4.3. *If the linear map $\phi: A \to A$ is defined by $\phi(1) = 0$ and $\phi(a(i_1 i_2 \dots i_n)) = a[i_1 i_2 \dots i_n]$, then $\phi(a(i_1 \dots i_r)a(i_{r+1} \dots i_{r+s})) = 0$ for all $i_1, i_2, \dots, i_{r+s} = 1, 2, \dots, m$; $r > 0$, $s > 0$.*

Proof. We shall proceed by induction on $n = r + s$. If $n = 2$, then the theorem can be verified easily. Assume $n > 2$ and that the theorem is proved for smaller values of n. By the definition of the multiplication in A, we have

$$\begin{aligned}
&\phi(a(i_1 \dots i_r)a(i_{r+1} \dots i_n)) \\
&\qquad = \sum \lambda(\sigma) a(i_{\sigma(1)}) \vee a[i_{\sigma(2)} \dots i_{\sigma(n)}] \\
&\qquad\quad - \sum \lambda(\sigma) \prod_{\nu=1}^{n-1} \lambda(i_{\sigma(n)} i_{\sigma(\nu)}) a(i_{\sigma(n)}) \vee a[i_{\sigma(1)} \dots i_{\sigma(n-1)}],
\end{aligned}$$

where the sums run over all shuffles of type (r, s), $r + s = n$. Since $\sigma(1) = 1$ or $r + 1$, and $\sigma(n) = r$ or n, the right-hand side of the above equation can be written

$$\begin{aligned}
&\sum_{\sigma(1)=1} \lambda(\sigma) a(i_1) \vee a[i_{\sigma(2)} \dots i_{\sigma(n)}] \\
&\qquad + \sum_{\sigma(1)=r+1} \lambda(\sigma) a(i_{r+1}) \vee a[i_{\sigma(2)} \dots i_{\sigma(n)}] \\
&\qquad - \sum_{\sigma(n)=r} \lambda(\sigma) \prod_{\nu=1}^{n-1} \lambda(i_r i_{\sigma(\nu)}) a(i_r) \vee a[i_{\sigma(1)} \dots i_{\sigma(n-1)}] \\
&\qquad - \sum_{\sigma(n)=n} \lambda(\sigma) \prod_{\nu=1}^{n-1} \lambda(i_n i_{\sigma(\nu)}) a(i_n) \vee a[i_{\sigma(1)} \dots i_{\sigma(n-1)}] \\
&= a(i_1) \vee \phi(a(i_2 \dots i_r)a(i_{r+1} \dots i_n)) \\
&\qquad + \prod_{\nu=1}^{r} \lambda(i_{r+1} i_\nu) a(i_{r+1}) \vee \phi(a(i_1 \dots i_r)a(i_{r+2} \dots i_n)) \\
&\qquad - \prod_{\mu=r+1}^{n} \lambda(i_\mu i_r) \prod_{\nu=1, \neq r}^{n} \lambda(i_r i_\nu) a(i_r) \vee \phi(a(i_1 \dots i_{r-1})a(i_{r+1} \dots i_n)) \\
&\qquad - \prod_{\nu=1}^{r-1} \lambda(i_n i_\nu) a(i_n) \vee \phi(a(i_1 \dots i_r)a(i_{r+1} \dots i_{n-1})) \\
&= 0
\end{aligned}$$

because of the induction assumption and the fact that, for $r = 1$,

$$\prod_{\mu=r+1}^{n} \lambda(i_\mu i_r) \prod_{\nu=1,\neq r}^{n} \lambda(i_r i_\nu) = 1.$$

COROLLARY 4.3. *If* $0 < r < n$, *then*

$$na(i_1, \ldots, i_r)a(i_{r+1} \ldots i_n) = \sum \lambda(\sigma)(na(i_{\sigma(1)} \ldots i_{\sigma(n)}) - a[i_{\sigma(1)} \ldots i_{\sigma(n)}]),$$

where the sum ranges over all shuffles of type $(r, n - r)$.

The above corollary, together with (4.2), shows that the $(n-1)m^n$ elements $a(i_1 \ldots i_r)a(i_{r+1} \ldots i_n)$, $i_1, \ldots, i_n = 1, 2, \ldots, m$; $0 < r < n$, and the m^n elements $na(i_1 \ldots i_n) - a[i_1 \ldots i_n]$ span the same vector space over K. Also from (4.2) we obtain

COROLLARY 4.4. *The linear map* $\phi_0 : A \to A$ *defined by* $\phi_0(1) = 0$ *and*

$$\phi_0(a(i_1 i_2 \ldots i_n)) = n^{-1}a[i_1 i_2 \ldots i_n],$$

for $n > 0$, *is a projection, that is,* $\phi_0^2 = \phi_0$.

The following theorem is essentially a generalization of Theorem 2.6 of **(2)**, and may be proved by using the map D introduced in the above.

THEOREM 4.5. *For* $n > 0$, *we have*

$$\sum_{s=0}^{n} (-1)^s \prod_{s<\mu<\nu<n} \lambda(i_\nu i_\mu)a(i_1 \ldots i_s)a(i_n i_{n-1} \ldots i_{s+1}) = 0.$$

5. Generalization of a theorem of Friedrichs. Let L be a free χ-aigebra of rank m, and $\eta : L \to U$ the linearization of L into its universal enveloping algebra. Let U' be a replica of U with the grade-respecting isomorphism $\iota : U \to U'$ and $\eta' = \iota \circ \eta$. Let $U \otimes U'$ be the tensor product of U and U' with respect to χ. In the course of the proof of (3.1) we have seen that the map $\bar{\eta} : L \to U \otimes U'$ defined by

$$\bar{\eta}(f) = \eta(f) \otimes 1 + 1 \otimes \eta'(f)$$

is a linearization and that there exists a homomorphism $\xi : U \to U \otimes U'$ such that $\xi \circ \eta = \bar{\eta}$. Now the following theorem generalizes a theorem of Friedrichs **(2)**.

THEOREM 5.1. *Let* η, ι, *and* ξ *be as above. Then an element* u *in* U *belongs to the image* $\eta(L)$ *of* L *under* η *if and only if*

$$\xi(u) = u \otimes 1 + 1 \otimes \iota(u).$$

Proof. The "only if" part follows from the fact that $\bar{\eta} = \xi \circ \eta$. In order to prove the "if" part, let $x_1, x_2, \ldots, x_m$ be free generators of U and write, for simplicity, x_i and x_i' for $x_i \otimes 1$ and $1 \otimes \iota(x_i)$, respectively. If

$$u = \sum \alpha_{i_1 \ldots i_n} x_{i_1} \ldots x_{i_n}$$

with coefficients in K, then

$$\xi(u) = \sum \alpha_{i_1 \dots i_n}(x_{i_1} + x'_{i_1}) \dots (x_{i_n} + x'_{i_n})$$

$$= \sum \sum_{s=0}^{n} \phi(a(i_1 \dots i_s)a(i_{s+1} \dots i_n))x_{i_1} \dots x_{i_s}x'_{i_{s+1}} \dots x'_{i_n},$$

where ϕ is a linear map: $A_n \to K$ defined by

$$\phi(a(i_1 \dots i_n)) = \alpha_{i_1 \dots i_n}.$$

Hence the condition given in (5.1) is equivalent to

$$\phi(a(i_1 \dots i_s)a(i_{s+1} \dots i_n)) = 0 \qquad (0 < s < n).$$

The rest of the proof is exactly the same as in the case $\lambda(ij) = 1$ **(2**, p. 214**)**, and may be omitted. Here we have to use

$$\sum a[i_1 \dots i_n]x_{i_1} \dots x_{i_n} = \sum a(i_1 \dots i_n)[x_{i_1} \dots x_{i_n}],$$

but this, too, can be proved as in **(2**, p. 213**)**.

Similarly we may prove the following

THEOREM 5.2. *A homogeneous element*

$$f = \sum \alpha_{i_1 \dots i_n}x_{i_1} \dots x_{i_n}$$

in U of degree $n > 0$ is a generalized Lie element if and only if

$$nf = \sum \alpha_{i_1 \dots i_n}[x_{i_1} \dots x_{i_n}].$$

This generalizes a theorem of Dynkin-Specht-Wever **(2**, p. 214**)**.

REFERENCES

1. C. Chevalley, *Fundamental concepts of algebra* (New York: Academic Press Inc., 1956).
2. Rimhak Ree, *Lie elements and an algebra associated with shuffles*, Ann. Math., *68* (1958), 210–220.
3. Seminaire "Sophus Lie," *Theorie des algebres de Lie et topologie des groupes de Lie*, 1954–5.
4. Ernst Witt, *Treue Darstellung Liescher Ringe*, J. Reine Angew. Math., *177* (1937), 152–160.

University of British Columbia

MODIFICATIONS AND COBOUNDING MANIFOLDS

ANDREW H. WALLACE

Introduction. The object of this paper is to establish a simple connection between Thom's theory of cobounding manifolds and the theory of modifications. The former theory is given in detail in **(8)** and sketched in **(3)**, while the latter is worked out in **(1)**. In particular in **(1)** it is shown that the only modifications which can transform one differentiable manifold into another are what I call below spherical modifications, which consist in taking out a sphere from the given manifold and replacing it by another. The main result is that manifolds cobound if and only if each is obtainable from the other by a finite sequence of spherical modifications.

The technique consists in approximating the manifolds by pieces of algebraic varieties. Thus if M_1 and M_2 form the boundary of M, the last is taken to be part of an algebraic variety such that M_1 and M_2 are two members of a pencil of hyperplane sections. If this pencil is properly chosen it will cut only finitely many singular sections on M, each of which will correspond to a spherical modification. The converse result is proved by a construction which seeks to bring about the situation just sketched. These results are proved in the first three sections.

The situation described here is essentially the same as arises in the study of critical values of a function on a manifold. Thus if M is embedded in N-space, each modification on the way from M_1 to M_2 corresponds to a critical value of x_N. The main result of § 4 is to show that the embedding can be done in such a way that, as x_N increases from its value on M_1 to its value on M_2, the type numbers of these critical points **(7**, p. 21**)** do not decrease. Whether the theory of critical points could be used more extensively in the present connection is not quite clear. One factor arising here (as for example in § 5) is that M_1 and M_2 are the main objects of interest usually, and the M which they cobound may be altered in some way, whereas the application of critical point theory would require that M should not be changed but should be treated as the underlying space. At any rate so far any application of, say, the Morse inequalities **(7**, p. 85**)** has yielded only trivial results.

Section 5 shows how the same effect may be brought about sometimes by modifications of different types, and the result is applied to give a solution of a problem of Bing **(2)** on the structure of 3-manifolds.

In § 6 it is shown that any differentiable manifold of dimension not less

Received January 19, 1959.

than 3 cobounds a simply connected manifold, while in § 7 a few results are given extending this to higher homology and homotopy groups.

1. Spherical modifications. Throughout this discussion E^n and S^n will denote an n-dimensional cell and an n-dimensional sphere, respectively, subscripts being used where necessary to distinguish between different copies of these sets.

Let M be a differentiable manifold of dimension n, and let S^m be an m-sphere homeomorphically and differentiably embedded in M. It is known that a sufficiently small neighbourhood B of S^m in M can be fibred by $(n-m)$-dimensional cells; B is then the normal bundle of S^m in M. If B can be expressed as the topological product $S^m \times E^{n-m}$, S^m will be said to be directly embedded in M. In this case the frontier of B, or what is the same thing, the frontier of $M-B$ is of the form $S^m \times S^{n-m-1}$. The last product can, however, be identified with the frontier of a product of the type $E^{m+1} \times S^{n-m-1}$. It follows at once that the union of $M-B$ and $E^{m+1} \times S^{n-m-1}$, corresponding points on the frontiers of these sets being identified, can be made into a differentiable manifold M'. The transition from M to M' is a modification **(1)**. Modifications constructed in this particular way from directly embedded spheres will be called spherical modifications. To draw attention to the dimensions involved, the modification from M to M' described above will be called a modification of type $(m, n-m-1)$; it can easily be seen that the inverse operation, going from M' to M, is a spherical modification of type $(n-m-1, m)$. It will also sometimes be convenient to describe the modification from M to M' as a modification which shrinks S^m and introduces S^{n-m-1}.

It is clear from the above description that the manifold M' contains a directly embedded sphere S^{n-m-1} and that $M - S^m$ and $M' - S^{n-m-1}$ are, in a natural way, homeomorphic. This homeomorphism will be said to be induced by the modification.

Still using the above notations, it is not hard to see that the result of a spherical modification does not depend essentially on the way in which B is fibred by cells transversal to S^m. This follows from the fact that every such fibring can be continuously deformed into a canonical fibring by cells made up from geodesic arcs normal to S^m, with respect to some Riemannian metric on M. Similarly, isotopic deformations of S^m will not affect the modifications. On the other hand the mode of expression of B as a product $S^m \times E^{n-m}$, equivalent to the choice of a system of cross-sections of B, may be an essential factor in determining the result of the modification. Thus it is not in general possible to speak of *the* modification shrinking S^m unless reference is also made to the way in which B is written as a product.

2. Cobounding manifolds. A differentiable manifold with boundary is a topological space M with a subspace M_1 such that (1) M_1 is a differentiable manifold; (2) each point of $M - M_1$ has a neighbourhood homeomorphic to

an n-cell (n the same for each point); (3) each point of M_1 has a neighbourhood in M homeomorphic to a solid n-dimensional hemisphere, the base of the hemisphere corresponding to the part of the neighbourhood on M_1; and (4) the transition functions between one neighbourhood and another of the types just described are differentiable. When M and M_1 are related in this way, M_1 will be said to be the boundary of M, and M_1 will be said to be a bounding manifold. In all this there is no need for the manifolds to be connected. Two differentiable manifolds M_1 and M_2 will be said to be cobounding if their union is a bounding manifold.

In the case of orientable manifolds the idea of bounding can be made a bit stronger. If M_1 is orientable it will be said to be an oriented bounding manifold if it is the boundary of an oriented manifold whose orientation induces a preassigned orientation of M_1. The set of all orientable manifolds is now taken as the set of generators of an additive abelian group. Each connected manifold is supposed to be given a preassigned orientation, and the minus sign denotes change of this orientation. The manifolds M_1 and M_2 are now said to be cobounding if $M_1 - M_2$ is an oriented bounding manifold.

From the algebraic point of view, the notion of cobounding introduced at the beginning of this section can be described as cobounding modulo 2.

The first main result to be proved is the following connection between the ideas of cobounding and of spherical modifications.

THEOREM 1. *Let M_1 and M_2 be two given compact differentiable manifolds, the question of orientation being for the moment ignored. Then M_1 and M_2 are cobounding if and only if each can be obtained from the other by a finite sequence of spherical modifications.*

Proof. The "if" part of the theorem will be established if it is shown that M_1 and M_2 cobound whenever one is obtained from the other by a single spherical modification, since the relation of cobounding is transitive. This will be proved now as part (a) of the proof, part (b) being the proof of the converse.

(a) Suppose then that M_1 is obtained from M_2 by a spherical modification of type $(r, n-r-1)$, n being the dimension of the manifolds. Thus there are spheres S^r and S^{n-r-1} contained respectively in M_1 and M_2, with normal bundles $B_1 = S^r \times E^{n-r}$ and $B_2 = E^{r+1} \times S^{n-r-1}$ in these manifolds such that $M_1 - B_1$ and $M_2 - B_2$ are homeomorphic. Assume now that $M_1 - B_1$ and $M_2 - B_2$ are identified with $(M_1 - B_1) \times \{0\}$ and $(M_1 - B_1) \times \{1\}$, respectively, in the set $(M_1 - B_1) \times I$, where I is the unit interval $0 \leqslant t \leqslant 1$. Form the union $[(M_1 - B_1) \times I] \cup B_1 \cup B_2$, B_1 and B_2 being inserted where they belong in $(M_1 - B_1) \times \{0\}$ and $(M_1 - B_1) \times \{1\}$ according to the identification just made. The subset $B_1 \cup B_2 \cup (\mathrm{Fr}B_1 \times I)$ in the space so constructed is an n-sphere and so can be identified with the boundary of an $(n+1)$-cell E^{n+1}. Adding E^{n+1} to $[(M_1 - B_1) \times I] \cup B_1 \cup B_2$ with suitable identifications on the boundaries, an $(n+1)$-dimensional manifold M is obtained, and can easily be adjusted along the boundary of E^{n+1} so as to be

differentiable. Moreover, it is clear that the boundary of M is the union of M_1 and M_2. Thus M_1 and M_2 are cobounding manifolds as was to be shown.

(b) The idea of the converse is as follows. Suppose M is a differentiable manifold with boundary, the boundary being the union of M_1 and M_2. It is to be shown that M_2 can be obtained from M_1 by a finite sequence of spherical modifications. To show this, M is first to be approximated by part of a real algebraic variety in N-space in such a way that M_1 and M_2 are parts of the sections by the hyperplanes $x_N = 0$ and $x_N = 1$, respectively. This can be done in such a way that the family of hyperplanes $x_N = c$, for $0 \leqslant c \leqslant 1$, cuts the approximation of M in non-singular sections with just a finite number of exceptions, on each of which there is exactly one singular point at which the tangent cone is a non-degenerate quadric cone. Then it will be shown that the transition from one side of a singular section to the other is locally the same as the transition from negative to positive values of t in a family of quadrics

$$\sum_{i=1}^{n} a_i x_i^2 = t$$

in n-space, and hence it will be verified that each such transition is carried out by means of a spherical modification.

The details of the proof just sketched will now be worked out. In the first place M is to be embedded in a Euclidean N-space E_N, which can be done if N is large enough. Also it is clear that the embedding can be done in such a way that M_1 and M_2 lie in the hyperplanes $x_N = 0$ and $x_N = 1$, respectively, while the rest of M lies entirely between these hyperplanes. The algebraic approximation mentioned above could be made already at this stage, but to ensure that the approximating variety will have no points near M except those which are actually approximating points of M it is convenient to carry out the following additional construction. Take second copies in E_N of M, M_1, M_2, respectively, namely M', M_1', M_2', and suppose that M_1' and M_2' lie in the hyperplanes $x_N = 0$ and $x_N = 1$, respectively, and that the rest of M' lies between these hyperplanes; also assume that $M \cap M' = \emptyset$. M' can be constructed in this way by a translation in E_N for example. In addition M and M' can be adjusted so that they cut the hyerplanes $x_N = 0$ and $x_N = 1$ orthogonally. By adding to $M \cup M'$ sets homeomorphic to $M_1 \times I$ and $M_2 \times I$, lying in the parts of E_N where $x_N \leqslant 0$ and $x_N \geqslant 1$, respectively, a compact differentiable manifold M'' can be constructed. M'' has the property that there is a neighbourhood U of M in E_N such that $U \cap M''$ is homeomorphic to M; in fact it is equal to M with, so to speak, a narrow fringe added along M_1 and M_2. Now it is known **(4; 9)** that there is a real algebraic variety V in E_N with an isolated sheet approximating M'' arbitrarily closely. This approximation is not only in the pointwise sense, but also the tangent linear varieties at corresponding points of M'' and V approximate one another arbitrarily closely **(4; 9)**. In particular it follows that M itself is approximated

arbitrarily closely by the part of $V \cap U$ which lies between $x_N = 0$ and $x_N = 1$, while M_1 and M_2 are approximated by the intersections of $V \cap U$ with these hyperplanes.

At this stage it is convenient to make a change of notation, simply replacing M by its approximation. Thus from now on in this proof it will be assumed that M lies on a real algebraic variety V in E_N and that there is a neighbourhood U of M such that M is the part of $V \cap U$ lying between the hyperplanes $x_N = 0$ and $x_N = 1$, while M_1 and M_2 are the intersections of these hyperplanes with M.

Some properties of an algebraic variety in relation to a pencil of hyperplane sections are now to be applied to the present situation. In the first place, if V is an algebraic variety in real projective space and Π is a generic hyperplane pencil only a finite number of members of Π will contain the tangent linear variety at some simple point of V, and each of these will contain the tangent linear variety at exactly one point of V. In addition, each of these finitely many points of contact for members of Π is a generic point of V. This can all be proved as in **(10**, ch. 1**)**. The fact that V may not be non-singular makes no essential difference to the technique of the dual variety used there. Now choose homogeneous co-ordinates $(x_1, x_2, \ldots, x_N, x_{N+1})$ in the space containing V such that $x_{N+1} = 1$ and the equations of the members of Π are of the form $x_N =$ constant, and also such that, if V is of dimension m, the projection of V into the linear subspace $x_{m+1} = x_{m+2} = \ldots = x_{N-1} = 0$ is one-one around a generic point. When this is done the equations of V (in affine form) will be

$$\left.\begin{aligned} F(x_1, x_2, \ldots, x_m, x_N) &= 0 \\ x_i &= R_i(x_1, x_2, \ldots, x_m, x_N) \end{aligned}\right\} \tag{1}$$

where $i = m + 1, m + 2, \ldots, N - 1$, F being a polynomial and the R_i rational functions with coefficients which are real when V is a real variety. Also, making a shift of origin to one of the points at which a member of Π contains the tangent linear space to V, and remembering that such a point is generic on V over the real numbers it turns out that equations (1) can be written in the form

$$\begin{aligned} x_N &= f(x_1, x_2, \ldots, x_m) \\ x_i &= g_i(x_1, x_2, \ldots, x_m) \end{aligned} \tag{2}$$

where $i = m + 1, m + 2, \ldots, N - 1$, and the functions f and the g_i are real analytic in a sufficiently small neighbourhood of the origin. Also, since the new origin started off as a generic point of V the power series expansion for f around that point is of the form

$$f = \sum_{i,j=1}^{m} a_{ij} x_i x_j + \ldots \tag{3}$$

where the dots denote terms of order greater than two and the determinant

$|a_{ij}|$ is not zero. The linear terms are of course zero because the tangent linear variety to V at the origin is contained in $x_N = 0$.

Now in what has just been said the pencil Π is generic, that is to say, the coefficients of the linear equations defining the axis of Π are indeterminates over the real numbers. The conditions that the choice of Π and of co-ordinates as above should not give equations for V of the type (2) and (3) at each of the points where a member of Π contains the tangent linear variety is algebraic in these indeterminates. It follows that the coefficients of the equations of the axis of Π can be given real values in such a way that the equations of V can be brought into the form described above. A final point is that, since the pencils which are unfavourable lie in an algebraic family, then whatever co-ordinate system is given in the space containing V, a linear change of co-ordinates with a matrix whose elements are arbitrarily close to those of the identity matrix will yield a co-ordinate system in which the equations of V can be written in the manner just described.

The discussion just carried out is now to be applied to the variety V of dimension $n + 1$ introduced in the earlier part of this proof, namely the real variety containing the manifold with boundary M whose sections with $x_N = 0$ and $x_N = 1$ are the manifolds M_1 and M_2 respectively. Then a small displacement of the given co-ordinate system will give a system with the following properties. There is a neighbourhood U of M such that the intersection of $U \cap V$ with $x_N = c$ is non-singular for all except a finite set of values $c_1, c_2, \ldots, c_k$ of c; for each i, $x_N = c_i$ intersects $U \cap V$ in a section with exactly one singular point, say P_i; if P_i is taken as origin the equations of V can be written in the form (2) and (3) around P_i. Since V was approximately orthogonal to $x_N = 0$ and $x_N = 1$ at points of M_1 and M_2 in terms of the original co-ordinates, and since the displacement of co-ordinates is supposed to be small, it follows that the intersections of $x_N = 0$ and $x_N = 1$ with $U \cap V$ in the new co-ordinates are respectively homeomorphic to M_1 and M_2. Again it is convenient to change the notation and simply to say that these intersections are M_1 and M_2.

To complete the proof of the theorem it will be shown that the transition from the intersection of $U \cap V$ with $x_N = c_i - \epsilon$ to its intersection with $x_N = c_i + \epsilon$, for some small positive ϵ can be made by means of a spherical modification. To do this fix attention on one of the P_i and take it as origin. Then in a neighbourhood of the origin V will have equations of the type (2), with f of the form (3). With this new arrangement of the co-ordinates the section $M(c)$ of M by the hyperplane $x_N = c$, for sufficiently small c, will have equations in a neighbourhood of the origin of the form

$$\sum a_{ij}x_i x_j + \phi = c \tag{4}$$

where ϕ is a power series in the variables $x_1, x_2, \ldots, x_{n+1}$ of order not less than three, along with further equations which express $x_{n+2}, x_{n+3}, \ldots, x_{N-1}$ as analytic functions of $x_1, x_2, \ldots, x_{n+1}$. By a linear change of the variables

$x_1, x_2, \ldots, x_{n+1}$ the quadratic terms in (4) can be diagonalized. Assuming that this has been done, (4) will be of the form

$$\sum_{i=1}^{n+1} a_i x_i^2 + \phi = c. \tag{5}$$

Since ϕ contains only terms of degree greater than two, a theorem of Samuel **(5)** shows that, for sufficiently small values of the variables, an analytic change of co-ordinates from $x_1, x_2, \ldots, x_{n+1}$ to a new set $y_1, y_2, \ldots, y_{n+1}$ can be made by formulae of the type $x_i = y_i + h_i(y)$, where the h_i are power series of order not less than two, in such a way that

$$\sum a_i x_i^2 + \phi = \sum a_i y_i^2.$$

By orthogonal projection from $(x_1, x_2, \ldots, x_N)$-space into $(x_1, x_2, \ldots, x_{n+1})$-space followed by a change to the y-co-ordinates it is then clear that a neighbourhood of the origin on V, that is to say on M, can be mapped analytically and homeomorphically on a neighbourhood of the origin in $(y_1, y_2, \ldots, y_{n+1})$-space, and the parts of the $M(c)$ near the origin in N-space will be mapped into the family of quadrics $Q(c)$, or at least the parts of these quadrics near the origin, in $(y_1, y_2, \ldots, y_{n+1})$-space, where $Q(c)$ has the equation

$$\sum_{i=1}^{n+1} a_i y_i^2 = c. \tag{6}$$

Now it can be explicitly verified that if $r+1$ of the a_i in (6) are positive and the rest negative (none are zero) and if c_0 is positive then the transition from $Q(c_0)$ to $Q(-c_0)$ can be made by a spherical modification of type $(r, n-r+1)$. In addition the homeomorphism induced by this modification can be constructed in a particular way. Namely, if small neighbourhoods, more precisely normal bundles, of the spheres

$$\sum_{i=1}^{r+1} a_i y_i^2 = c_0,\ y_j = 0 \qquad (j > r+2)$$

on $Q(c_0)$ and

$$\sum_{i=r+2}^{n+1} a_i y_i^2 = -c_0,\ y_j = 0 \qquad (j < r+1)$$

on $Q(-c_0)$ are removed (here it is assumed that $a_1, a_2, \ldots, a_{r+1}$ are the positive a_i) then the corresponding points on the remaining sets of $Q(c_0)$ and $Q(-c_0)$ are joined to each other by members of the family F of orthogonal trajectories to the family of $Q(c)$.

Returning to the variety V and more specifically to M, it has already been seen that $y_1, y_2, \ldots, y_{n+1}$ can be taken as a set of local analytic co-ordinates on M around the origin. Also the ordinary Euclidean metric in $(y_1, y_2, \ldots, y_{n+1})$-space induces a Riemannian metric on M in a neighbourhood of the origin. By means of a partition of unity a Riemannian metric can be set up on the whole of M so as to agree with this induced metric in a sufficiently

small neighbourhood of the origin on M. Then the image on M of the family F of orthogonal trajectories to the $Q(c)$ can be extended to the family F' of orthogonal trajectories to the family of sections $M(c)$ of M, at least in a neighbourhood of $M(0)$. It is thus clear that, for c_0 sufficiently small and positive, if the images on M of the spheres on $Q(c_0)$ and $Q(-c_0)$ mentioned above are removed, then the remaining sets on $M(c_0)$ and $M(-c_0)$ are homeomorphic, corresponding points being joined by members of the family F'. Apart from this the spherical modification carrying $Q(c_0)$ into $Q(-c_0)$, in so far as it affects points near the origin, is carried into a similar modification taking $M(c_0)$ into $M(-c_0)$. And this completes the proof of the theorem.

It is possible to give part (a) of the above theorem a more precise form. Namely, if M_2 is obtained from M_1 by a single spherical modification, then the manifold M can be constructed in such a way that M_1 and M_2 belong to a pencil of hyperplane sections of M containing exactly one singular section. In other words the given modification can be made to arise in the same way as the modifications shown to exist in part (b) of the theorem. To prove this, the cell E^{n+1} which appeared in the course of the proof of part (a) must be constructed in a special way. For values of t such that $-1 \leqslant t \leqslant 1$, let $Q(t)$ be the quadric hypersurface $x_1^2 + x_2^2 + \ldots + x_{r+1}^2 - x_{r+2}^2 - \ldots - x_{n+1}^2 = t$ in $(n+1)$-space. The section of $Q(1)$ by the linear space $x_{r+2} = x_{r+3} = \ldots = x_{n+1} = 0$ is an r-sphere S^r whose normal bundle of some convenient radius in $Q(1)$ is a set B_1' homeomorphic to $S^r \times E^{n-r}$, and so to B_1 (in the notation of part (a) of the above theorem). Construct the family of orthogonal trajectories F to the family $Q(t)$. Then the set of points on curves of F meeting $Q(1)$ at points of B_1' is an $(n+1)$-cell $E'^{(n+1)}$. It is clear that, apart from the curves of F starting at points of S^r, all of which end at the origin, all the members of F starting at points of B_1' reach $Q(-1)$ at points in the normal bundle B_2' of the sphere S^{n-r-1} in which $Q(-1)$ is cut by the linear space $x_1 = x_2 = \ldots = x_{r+1} = 0$, and similarly the other way round. B_2' is homeomorphic to $S^{n-r-1} \times E^{r+1}$, that is to say, to B_2. Now, referring to the proof of part (a) of the above theorem, it will be seen that the frontier of E^{n+1} first appeared as the frontier of $(M_1 - B_1) \times I$ with the sets B_1 and B_2 added in the appropriate way, M_1 and M_2 being identified with $(M_1 - B_1) \times \{0\} \cup B_1$ and $(M_1 - B_1) \times \{1\} \cup B_2$, respectively. The frontiers of E^{n+1} and $E'^{(n+1)}$ are now to be identified. To do this define a mapping f of the frontier of $E'^{(n+1)}$ onto that of E^{n+1} as follows: first f is to be defined as a homeomorphism of B_1' onto B_1 preserving the product structure. Then if (p, t) is the point of parameter t (that is, the point lying on $Q(t)$) on the curve of the family F which passes through p on B_1', $f(p, t)$ will be defined as the point $(f(p), \frac{1}{2} - \frac{1}{2}t)$ in $\mathrm{Fr}B_1 \times I$ (this makes sense as $f(p)$ is already defined). In particular f is now defined as a homeomorphism of $\mathrm{Fr}B_2'$ onto $\mathrm{Fr}B_2$, preserving the product structure, and so it can be extended over the whole of B_2', carrying this set homeomorphically onto B_2. f is now defined on the whole of $\mathrm{Fr}E'^{(n+1)}$, and so can be extended to a homeomorphism of $E'^{(n+1)}$ onto $E^{(n+1)}$.

Using the mapping f just defined, the family $M(t)$ of sets will now be defined. For each t such that $-1 \leqslant t \leqslant 1$ set

$$M(t) = f(Q(t) \cap E'^{(n+1)}) \cup (M_1 - B_1) \times \{s\}$$

where $s = \frac{1}{2} - \frac{1}{2}t$. Then, for each $t \neq 0$, $M(t)$ is a manifold, and $M(0)$ has a single isolated singular point corresponding to the vertex of the cone $Q(0)$. In particular $M(1) = M_1$ and $M(-1) = M_2$.

If the family $M(t)$ is in $(x_1, x_2, \ldots, x_{N-1})$-space, then the manifold M of part (a) of Theorem 1 can be constructed in $(x_1, x_2, \ldots, x_N)$-space by taking it as the set whose intersection with $x_N = t$ is $M(t)$. As the construction has been done here, M and the $M(t)$ may not be differentiable, but they can clearly be arranged to be so by taking suitable precautions when the boundary of E^{n+1} and that of $(M_1 - B_1) \cup B_1 \cup B_2$ are identified, and when the mapping f is extended into the interior of $E'^{(n+1)}$.

A further point to notice is the existence of a family F of curves on M consisting of the image under f of the orthogonal trajectories to the $Q(t)$ lying in $E'^{(n+1)}$ along with all the curves of the form $\{p\} \times I$ for p in $M_1 - B_1$. These curves have the following properties:

(1) Exactly one of them passes through each point of M different from P, the image under f of the origin in $(x_1, x_2, \ldots, x_{n+1})$-space.

(2) The curves starting on S^r in M_1 all end at P; so also do those which start at points of S^{n-r-1} in M_2.

(3) The set of points on the members of F starting on S^r is an $(r+1)$-cell E^{r+1} in M. Thus E^{r+1} is an $(r+1)$-cell in M with boundary S^r on M_1. Similarly there is an $(n-r)$-cell E^{n-r} in M with boundary S^{n-r-1} on M_2.

Suppose that, in addition to the modification ϕ carrying M_1 into M_2, a second modification ϕ' is applied to M_2, taking it into M_3, and suppose that a manifold M' having M_2 and M_3 as its boundary and containing a family F' of curves with properties similar to (1), (2), and (3) above has been constructed in the manner just described for M and F. Then M and M' can be joined together along M_2, and if suitable precautions are taken the result will be a differentiable manifold. Also the families F and F' can be combined, each curve of F being joined to the curve of F' starting at its end point on M_2. Now it has been remarked that a displacement of the sphere shrunk in a modification does not affect the result, and so if ϕ' is of type $(s, n-s-1)$ with $s \leqslant r$ it can always be arranged that the S^s shrunk by ϕ' does not meet the S^{n-r-1} introduced by ϕ. It follows that the curves of F' starting on S^{n-r-1} can be added to E^{n-r} to give a larger $(n-r)$-cell in $M \cup M'$ with its boundary in M_3. A similar remark can be made concerning any sequence of modifications of suitable types.

It should be remarked here that, in the proof of part (b) of the above theorem, there is an extreme case which may occur, corresponding to the values -1 or n for r. This arises when a section $x_N = c$ of M has a singularity which is an isolated point. Although, strictly speaking this should be allowed as

a modification with the appropriate alteration to the statement of Theorem 1, it will turn out (cf. § 4, Theorem 4) that these extreme cases can be avoided by suitably transforming the manifold M.

3. The oriented case. For the present purpose the most convenient way of fixing the orientation of a connected orientable differentiable manifold is by means of sets of local co-ordinates. Namely, having fixed a co-ordinate system in a neighbourhood U, a second system in U will be called positively or negatively oriented according as the Jacobian of the co-ordinate transformation is positive or negative. For a connected orientable manifold there is a covering by co-ordinate neighbourhoods with co-ordinates chosen so that, in the overlap of any two of the neighbourhoods the Jacobian of the corresponding co-ordinate transformation is positive. If the restriction to U of any one of these co-ordinate systems is positively oriented then the whole collection of local co-ordinate systems defines on the manifold the orientation induced by the fixed system in U.

The following lemmas prepare the way for the main result of this section.

LEMMA 3.1. *Let M be a connected orientable differentiable manifold in Euclidean N-space, and let H be a hyperplane such that $H \cap M$ is a connected differentiable manifold. Then $H \cap M$ is orientable.*

Proof. Local co-ordinates can be taken on M in a neighbourhood U of a point of $H \cap M$ in such a way that, if the Euclidean co-ordinates have been arranged so that H has the equation $x_N = 0$, then x_N is one of the local co-ordinates. It is clear then that x_N can be included among the local co-ordinates around every point of $H \cap M$, and so the orientation induced on M by the selected co-ordinate system in U automatically defines an orientation on $H \cap M$, which is therefore orientable.

COROLLARY. *If M is an orientable differentiable manifold with a connected boundary which is also a differentiable manifold, then this boundary is also orientable.*

Proof. For the given manifold can be so arranged that the boundary is a hyperplane section.

In the above lemma it should be noted (and this observation also applies to the corollary) that, if $H \cap M$ is not connected, orientability holds for each of the connected components separately.

LEMMA 3.2. *Let M and M' be connected orientable differentiable manifolds having a common boundary which is a connected differentiable manifold M_1. Then $M \cup M'$ is orientable.*

Proof. Embed M and M' in N-space so that M is in the set $x_N \leqslant 0$ and M' in the set $x_N \geqslant 0$, M_1 thus being the section of $M \cup M'$ by $x_N = 0$. It is then easy to see that local co-ordinates in $M \cup M'$ can be chosen around

each point of M_1 so that x_N is always included as one of the co-ordinates, while the rest of $M \cup M'$ can be covered by co-ordinate neighbourhoods in M and M' separately. Since M and M' are orientable and M_1 is connected it follows at once that the co-ordinates can be chosen in each of these neighbourhoods so that an orientation is defined on $M \cup M'$ as required.

LEMMA 3.3. *Let M_1 be a connected orientable differentiable n-manifold, and let M_2 be obtained from M_1 by a spherical modification of type $(r, n-r-1)$ with r not equal to 0 or $n-1$. Then M_2 is orientable.*

Proof. Suppose that the modification in question shrinks the sphere S^r with normal bundle B_1 in M_1. Then $M_1 - B_1$ isan oriented manifold with a connected boundary. Also B_2 (the set to be added to $M_1 - B_1$ in the modification) is oriented with the same connected boundary. Then by Lemma 3.2 $M_2 = (M_1 - B_1) \cup B_2$ is orientable.

The condition on r in the last lemma cannot be dropped. For it is possible for a $(0, n-1)$- or $(n-1, 0)$-modification to change the orientability or otherwise of a manifold, as, for example, in the case of a (0,1)-modification applied to the surface of a sphere to make it into a Klein surface. Of course there are two ways in which a (0, 1)-modification can be applied to a sphere, the one giving a torus and the other a Klein surface. A similar situation holds in general. For the effect of a $(0, n-1)$-modification on a manifold M_1 is to remove two disjoint n-cells from M_1 (namely the normal bundle of the S^0 to be shrunk) and to identify the points of the two $(n-1)$-spheres which are their boundaries. Clearly there are essentially two different ways of making this identification, and if M_1 is orientable one of these ways will give an orientable M_2 and the other a non-orientable one. If the $(0, n-1)$-modification carries an orientable manifold into another orientable manifold, then the modification itself will be said to be orientable.

The following theorem now gives the necessary complement to Theorem 1 for the case of orientable manifolds.

THEOREM 2. *Let M_1 and M_2 be two orientable differentiable manifolds. Then, with suitable orientations of their connected components, they cobound in the oriented sense if and only if they are related by a finite sequence of spherical modifications of which each modification of type $(0, n-1)$ or $(n-1, 0)$ is orientable.*

Proof. If M_1 and M_2 cobound in the oriented sense, then, by definition, their union constitutes the boundary of an orientable manifold M, and the orientations of the various components of M_1 and M_2 are supposed to be those induced by some selected orientation of M. As in Theorem 1, M is to be taken as part of a real algebraic variety in N-space such that M_1 and M_2 are the sections of M by the hyperplanes $x_N = 0$ and $x_N = 1$, while the rest of M lies between these hyperplanes. Also just a finite number of the hyperplanes $x_N = c$ are to cut M in singular sections, each with exactly one singular

point as in Theorem 1. By Lemma 3.1 and the remark following it, each hyperplane $x_N = c$, except those cutting singular sections, cuts M in a differentiable manifold whose components are orientable,with orientations induced by that of M. It follows at once, by considering sections on either side of a singular section corresponding to a $(0, n-1)$- or $(n-1, 0)$-modification that each such modification must be orientable (noting that this terminology makes sense whether the modification affects one component only or has the effect of joining two components together, for these components all have well defined orientations). This completes the proof in one direction.

To prove the converse, let M_2 be obtained from M_1 by a sequence of spherical modifications in which each of type $(0, n-1)$ or $(n-1, 0)$ is orientable. Here it is assumed that the components of M_1 are given preassigned orientations. Then Lemma 3.3 along with the assumed orientability of the $(0, n-1)$- and $(n-1, 0)$-modifications ensures that, as each modification is performed, the result is orientable with a naturally induced orientation on each component. The final result is supposed to be M_2 with suitable orientations on its components. The object now is to show that M, constructed as in Theorem 1, part (a), is orientable, and that it can be oriented in such a way that the correct orientations are induced on the components of M_1 and M_2. Clearly it is sufficient to carry out the proof in the case where M_1 and M_2 are related by one spherical modification.

Consider then the construction of M in the proof of Theorem 1, part (a). If the modification in question is of type $(r, n-r-1)$ with r not 0 or $n-1$ it may as well be assumed that M_1 is connected, since such a modification will affect just one component. Then, in the notation of part (a) of Theorem 1, $(M_1 - B_1) \times I$ is orientable and it is not hard to see that its frontier along with B_1 and B_2 will make up an oriented S^n, the orientation induced by that of $(M_1 - B_1) \times I$. It follows at once that when the cell E^{n+1} is added to form M the latter will be orientable and its orientation will induce that of M_1 and M_2. In the case of a $(0, n-1)$-modification, assumed orientable, this assumption turns out to be exactly what is wanted to ensure that $(\mathrm{Fr}B_1 \times I) \cup B_1 \cup B_2$ will be an oriented n-sphere, M_1 and M_2 having been suitably oriented. Then as before the addition of an $(n+1)$-cell gives an orientable manifold as required.

4. Rearrangement of modifications. In general there is no guarantee that the members of a sequence of modifications can be commuted among themselves, for the spheres introduced by the earlier modifications may intersect those to be shrunk in the later ones and it may be impossible to disentangle them. There are, however, certain ways in which the order of a sequence of modifications can be changed, and these will be examined in this section.

THEOREM 3. *Let M_1 and M_2 be n-dimensional differential manifolds related by a sequence of spherical modifications of types $(n-p-1, p)$ for various values of p not less than r. Then the order of these modifications can be changed*

in such a way that all the $(n - r - 1, r)$-modifications are done last (M_1 being counted as the initial state).

Proof. The assumption on p is vacuous if r is zero, but otherwise the proof in this case is the same. The situation of § 2 will be assumed to hold here, in particular as described in the remarks at the end of the section following the proof of Theorem 1. Namely, M_1 and M_2 will be assumed to form the boundary of a differentiable manifold M in N-space, and in fact to be the sections of M by the hyperplanes $x_N = 0$ and 1, the rest of M lying between these. And among the sections of M by the family $x_N = c$ there are to be finitely many with a singularity, each corresponding to a spherical modification. It will also be assumed for the moment that none of these singular sections has an isolated point corresponding to an n-sphere which shrinks to a point and vanishes as the section $x_N = c$ varies from $c = 0$ to $c = 1$. This restriction will be removed later (cf. Theorem 4). The section of M by $x_N = t$ is to be denoted by $M(t)$, and as in § 2 there is to be a family F of curves in M cutting across the non-singular $M(t)$ transversally.

Starting from M_1 let ϕ be the first modification of type $(n - r - 1, r)$, corresponding to a section $M(c)$ of M with a singularity at the point P. Then, as remarked in § 2, it can be assumed that the spheres shrunk in later modifications do not meet the members of F which meet the r-sphere introduced by ϕ, since all other modifications are of type $(n - p - 1, p)$ with $p \geqslant r$. The points of all the curves of F starting at P and lying in the part of M for which $x_N \geqslant c$ form an $(r + 1)$-cell E^{r+1} with boundary S^r contained in M_2. The idea of this proof is to deform the family $M(t)$ in a neighbourhood of E^{r+1}, so obtaining a new family of submanifolds, some with singularities. M is then to be deformed so that this new family becomes a pencil of hyperplane sections, a finite number being singular. These singular sections will correspond to a sequence of spherical modifications leading from M_1 to M_2, and it will turn out that the modifications are all the same as those in the given sequence, but that ϕ now appears last.

The details of the idea just sketched will now be filled in. There is a neighbourhood U of P on M which is the homeomorphic image, under a mapping f, of a neighbourhood of the origin on the quadric Q in $(n + 2)$-space with the equation

$$z = y_1^2 + y_2^2 + \ldots + y_{r+1}^2 - y_{r+2}^2 - \ldots - y_{n+1}^2.$$

By means of this mapping the section $M(c + t)$ of M is locally identified with the section $Q(t)$ of Q given by $z = t$ (cf. the end of § 2, with the appropriate changes of notation). Also under this homeomorphism f the sphere introduced by the modification ϕ is the image of the sphere on Q given by $y_1^2 + y_2^2 + \ldots + y_{r+1}^2 = z$, $y_{r+2} = 0, \ldots, y_{n+1} = 0$, for some sufficiently small $z > 0$, and the family F restricted to U is the image of the family of orthogonal trajectories to the $Q(t)$ in a neighbourhood of the origin.

The next step is to construct a neighbourhood in M of the set E^{r+1} in a rather special way. First, in the neighbourhood U, take the smaller neighbourhood $f(U_0)$, image under f of the set in Q defined by the inequalities

$$|z| \leqslant \epsilon, y_{r+2}^2 + y_{r+3}^2 + \ldots + y_{n+1}^2 \leqslant \delta$$

for sufficiently small positive ϵ and δ. It is not hard to see that U_0 is an $(n+1)$-cell with boundary consisting of the following three sets:

(1) The part of $z = -\epsilon$ on Q such that

$$\sum_{r+2}^{n+1} y_i^2 \leqslant \delta.$$

(2) The set $|z| \leqslant \epsilon$ satisfying

$$\sum_{r+2}^{n+1} y_i^2 = \delta.$$

This is homeomorphic to $S^r \times S^{n-r-1} \times I$.

(3) The part of $z = \epsilon$ on Q with

$$\sum_{r+2}^{n+1} y_i^2 \leqslant \delta.$$

This is homeomorphic to $S^r \times E^{n-r}$.

The image of the set (3) under f is a neighbourhood B_0 of S_0^r in $M(c+\epsilon)$, S_0^r being the sphere introduced by ϕ. If B_0 is small enough all the curves of F meeting it can be continued up to M_2; let B_1 be the set of points on all these curves. Then define B as the union of B_1 and $f(U_0)$. Clearly B is a neighbourhood in M of E^{r+1} and is an $(n+1)$-cell with boundary consisting of the sets:

(1)′ The image under f of the set (1) above.

(2)′ The union of the image under f of (2) above with the set of points on curves of F meeting $\mathrm{Fr}B_0$ on $M(c+\epsilon)$.

(3)′ $B \cap M$.

Note that the set (2)′, like (2), is homeomorphic to $S^r \times S^{n-r-1} \times I$. In (2) I is identified with the interval $|z| \leqslant \epsilon$, and in (2)′ with the interval $c - \epsilon \leqslant t \leqslant 1$, t being the parameter specifying the sections $M(t)$.

The switching of the order of modifications so that ϕ comes last is carried out by constructing a new mapping g of U_0 in Q into M, this time mapping it onto the whole of B. This mapping will be defined by identifying the sets (1), (2), and (3) on the frontier of U with the sets (1)′, (2)′, and (3)′ on the frontier of B, and then extending into the interiors of these sets.

The mapping $g : \mathrm{Fr}U_0 \to \mathrm{Fr}B$ is defined as follows:

(a) The restriction of g to the set (1) is to coincide with f.

(b) g is to map (2) onto (2)′. It has been noted that both sets are homeomorphic, and in a natural way, to $S^r \times S^{n-r-1} \times I$. g will be defined by giving

a homeomorphism h of the interval I in (2), namely $-\epsilon \leqslant z \leqslant \epsilon$, and the interval I in (2)′, namely $c-\epsilon \leqslant t \leqslant 1$. h is to be defined in such a way that the interval $-\epsilon \leqslant z \leqslant 0$ is mapped on the interval $c-\epsilon \leqslant t \leqslant 1-\eta$, where η is chosen so that all the sections $M(t)$ with $t \geqslant 1-\eta$ are homeomorphic to M_2. Apart from this condition h can be arbitrary.

(c) g as defined in (a) and (b) is to be extended in the obvious way to map the set (3) on the set (3)′.

Finally, since U_0 and B are $(n+1)$-cells, g can be extended into the interior of U_0 to give a homeomorphism of U_0 onto B.

To define a new sequence of modifications relating M_1 and M_2, construct a family $M'(t)$ of subsets of M as follows:

For $t \leqslant c-\epsilon$, $M'(t) = M(t)$.

For $c-\epsilon \leqslant t \leqslant 1-\eta$, $M'(t)$ is the union of the part of $M(t)$ outside B with $g(Q(h^{-1}(t)) \cap U_0)$.

For $t \geqslant 1-\eta$, $M'(t) = M(t)$.

The $M'(t)$ as so defined may not be differentiable but can be made so (apart from a finite number each of which will have one singularity) by a suitable adjustment, or by a suitable definition of g in the first place. Define M' to be the set in $(x_1, x_2, \ldots, x_{N+1})$-space such that $M(t)$ is the section by $x_{N+1} = t$ (M of course is supposed to be in $(x_1, x_2, \ldots, x_N)$-space). In particular $M'(0) = M_1$ and $M'(1) = M_2$, and so M_1 and M_2 cobound the new manifold M', which, incidentally, is clearly homeomorphic to M.

Consider now the set of modifications corresponding to the singular members of the family $M'(t)$. For $t < 1-\eta$, the only singular $M'(t)$s are those corresponding to all the original modifications relating M_1 and M_2 except ϕ. $M'(1-\eta)$ is a singular section of M' corresponding to a $(n-r-1, r)$-modification ϕ'. And there are no further modifications.

ϕ' can be thought of as the modification ϕ shifted to the end of the sequence of modifications. To complete the proof of the theorem, each $(n-r-1, r)$-modification is to be shifted to the end in this way, and this can be done in a finite number of steps as above.

There are a number of remarks and corollaries connected with the theorem just proved. In the first place it must be emphasized that M', as constructed in the course of the proof, is homeomorphic to M; this point is of importance in certain applications where the main object of interest is not the pair of manifolds M_1 and M_2 but the manifold M which they bound. Another point is that ϕ was taken as the first $(n-r-1, r)$-modification starting from M_1. It is quite clear however that M_1 and M_2 could be replaced by two intermediate sections M_1' and M_2' of M, when the same method of proof would show that any $(n-r-1, r)$-modification can be moved to any later stage in the sequence of modifications leading from M_1 to M_2.

An essential result which must now be obtained is the possibility of removing the restriction imposed in Theorem 1, that no section of M by a hyperplane $x_N = c$ should have an isolated point.

THEOREM 4. *Let M_1 and M_2 cobound M, these manifolds being arranged as in Theorem* 1 *in Euclidean N-space, singular sections by hyperplanes $x_N = c$ corresponding to spherical modifications leading from M_1 to M_2. Then the embedding of M can be done in such a way that no section by a hyperplane $x_N = c$ has an isolated point.*

Proof. Proceeding from M_1 to M_2 let $M(c)$ (notation of Theorem 1) be the last section of M with an isolated point P corresponding to a vanishing sphere. That is to say $M(c)$ has the isolated point P and for small ϵ $M(c-\epsilon)$ has a small isolated sphere near P, while $M(c+\epsilon)$ has no points near P. Varying t from c downwards, the n-sphere introduced at P becomes joined to some other component of a section of M by a $(0, n-1)$-modification (possibly after some modifications have been applied to the sphere itself). Let ϕ be the inverse of this $(0, n-1)$-modification, corresponding to a singular section $M(c')$ of M, and then, for a sufficiently small ϵ, apply Theorem 3 to the part of M between $M(c'-\epsilon)$ and $M(c-\epsilon)$. The result is that it can be assumed that, in the sequence of modifications leading from M_1 to M_2, the last modification before the vanishing of the n-sphere at P is an $(n-1, 0)$-modification which isolates that sphere. This modification will still be called ϕ, and the corresponding singular section of M will be $M(c')$.

For t near c' but less than it, $M(t)$ contains an $(n-1)$-sphere $S^{n-1}(t)$ which is to be shrunk by the modification ϕ. The part of $M(t)$ on one side of $S^{n-1}(t)$ is an n-cell $E^n(t)$. As t tends to c', $E^n(t)$ closes up to form an n-sphere, and $M(t)$, for $c' \leqslant t < c$, contains this detached sphere $S^n(t)$ which shrinks to a point as t tends to c. It is clear that, for a sufficiently small positive ϵ, the union of all the $E^n(t)$ for $c'-\epsilon \leqslant t < c'$ and all the $S^n(t)$ for $c' \leqslant t \leqslant c$ is an $(n+1)$-cell E^{n+1}, having on its boundary the n-cell E^n formed by the union of all the $S^{n-1}(t)$ for $c'-\epsilon \leqslant t \leqslant c'$ ($S^{n-1}(t)$ reduces to a point for $t = c'$). E^{n+1} is homeomorphic to a solid $(n+1)$-dimensional hemisphere, E^n corresponding to the solid n-sphere forming the base, and so, corresponding to the fibring of the hemisphere by concentric n-dimensional hemispheres, E^{n+1} can be fibred by a family of n-cells $E_1^n(t)$ such that $S^{n-1}(t)$ is the frontier of $E_1^n(t)$.

Now define the family $M'(t)$ of subsets of M as follows:

$$M'(t) = M(t) \text{ for } t \leqslant c'-\epsilon;$$

$$M'(t) = (M(t) - E^n(T)) \cup E_1^n(t) \text{ for } c'-\epsilon \leqslant t < c';$$

$$M'(t) = M(t) - S^n(t) \text{ for } c' \leqslant t \leqslant c;$$

$$M'(t) = M(t) \text{ for } t > c.$$

Having done this, let M' be the set in $(N+1)$-space such that $M'(t)$ is its section by the hyperplane $x_{N+1} = t$. It is clear that M' can be adjusted to become a differentiable manifold, and that M_1 and M_2 will form its boundary. The singular sections of M' by members of the pencil $x_{N+1} = c$ correspond to

a sequence of spherical modifications leading from M_1 to M_2. These modifications are the same as the original ones (corresponding to the singular sections of M) with the exception that ϕ has now dropped out, and the section corresponding to the isolated point at P is no longer there. By means of a finite number of steps as just described, all singular sections with isolated points can be removed.

In connection with the proof of this theorem it should be noted that the manifold M' is homeomorphic to M.

The results of Theorems 3 and 4 can now be combined to give a stronger form of Theorem 1.

THEOREM 5. *Let the n-dimensional differentiable manifolds M_1 and M_2 form the boundary of the differentiable manifold M. Then M can be embedded in N-space, for sufficiently large N, as part of a real algebraic variety, M lying entirely between the hyperplanes $x_N = 0$ and $x_N = 1$. Only a finite number of sections by hyperplanes $x_N = c$ $(0 \leqslant c \leqslant 1)$ will have singular points, one point on each such section, and none of these singular points will be an isolated point of the section in question. Finally the embedding can be arranged in such a way that, in the sequence of modifications leading from M_1 to M_2, corresponding to the singular sections of M, all the $(r, n-r-1)$-modifications come before the $(s, n-s-1)$-modifications for each pair of integers r, s with $r < s$.*

Proof. The first part of the theorem is simply Theorem 1. The absence of isolated points on the singular sections of M can be brought about by Theorem 4, and the ordering of the modifications according to type can be done by repeated application of Theorem 3.

A further point to notice in connection with the last theorem is that the modifications of any one type can be rearranged freely among themselves. For consider the modifications of type $(r, n-r-1)$ with $2r \leqslant n$ (this inequality imposes no restriction since in the contrary case one can look at the sequence of modifications the other way round, starting from M_2). Repeated application of Theorem 3 will rearrange these modifications in any preassigned way. The question of identifying the modifications as they are permuted is settled by noting that, since $2r \leqslant n$, there is a set of disjoint r-spheres each to be shrunk by one of the modifications, and the modifications can be named according to the sphere shrunk.

Theorem 5 is a generalization of a well-known result concerning orientable 3-manifolds. Let M be an orientable 3-manifold with boundary formed by M_1 and M_2, arranged as in Theorem 5; no generality is lost here since a 3-manifold can be triangulated and then smoothed to give a differentiable manifold. The only modifications leading from M_1 to M_2 as in Theorem 5 will be of types $(0, 1)$ and $(1, 0)$, all those of the former type being done first. If now M is a closed manifold, it can be assumed to be contained between the hyperplanes $x_N = 1 + \epsilon$ and $x_N = -\epsilon$, for small positive ϵ, while the sections of M by the hyperplanes $x_N = 1$ and $x_N = 0$ will be 2-spheres M_2

and M_1, boundaries of 3-cells E_2 and E_1 which lie respectively in the sets $1 \leqslant x_N \leqslant 1 + \epsilon$ and $-\epsilon \leqslant x_N \leqslant 0$. Theorem 5 then implies that M is obtained by applying $(0,1)$-modifications to the surfaces of E_1 and E_2, filling the surfaces in as one goes to obtain two solids, whose boundaries are then identified. Since M is orientable, all the $(0,1)$-modifications are of orientable type (Theorem 2), and so the solids obtained are solid spheres with handles. That is to say the manifold M is constructed by taking the union of two solid handled spheres (necessarily of the same genus) and identifying their boundaries **(6**, p. 219**)**.

Clearly Theorem 5 gives a similar way of constructing a non-orientable 3-manifold. In this case, however, at least one of the modifications applied to the surfaces M_1 and M_2 must be of non-orientable type. Thus the two solids which are to be put together to form M must each have at least one handle twisted (in the manner of the Klein surface).

To formulate Theorem 5 as a generalization of this classical result on 3-manifolds, a generalized handled sphere can be defined as an $(n+1)$-dimensional solid obtained from a solid $(n+1)$-sphere by applying to its surface $(r, n-r-1)$-modifications, with $r \leqslant n-r-1$, filling out the surface at each stage to form an $(n+1)$-solid. Then Theorem 5 implies that any differentiable $(n+1)$-manifold can be expressed as the union of two generalized handled spheres with boundaries identified. In particular if M is orientable, all the $(0, n-1)$-modifications involved will be of orientable type.

5. Complementary modifications. Let M_1 be a differentiable n-manifold and let S^r be a directly embedded r-sphere to be shrunk by a spherical modification ϕ. Suppose also that S^r is the boundary of an $(r+1)$-cell non-singularly and differentiably embedded in M_1. When $B_1 = S^r \times E^{n-r}$ is removed from M_1, the remaining set will contain an $(r+1)$-cell ${E_1}^{r+1}$ with boundary $S^r \times \{p\}$ for some $p \in S^{n-r-1} = \mathrm{Fr}E^{n-r}$. When $B_2 = {E_2}^{r+1} \times S^{n-r-1}$ is added to make the modification ϕ, ${E_2}^{r+1}$ joins up with ${E_1}^{r+1}$ to form a sphere S^{r+1} in M_2. S^{r+1} is not necessarily directly embedded in M_2, but a sufficient condition for direct embedding is that the natural (the precise meaning of this overworked word in this context is explained below) product structure of the normal bundle of E^{r+1} in M_1 should induce the product structure on B_1 associated with the modification ϕ. If this condition is satisfied, a second modification ϕ' can be performed, shrinking S^{r+1} and transforming M_2 into a manifold M_3.

LEMMA 5.1. *Under the conditions just described M_1 and M_3 are homeomorphic.*

Proof. A normal neighbourhood (union of normal geodesic elements) of E^{r+1} in M_1 is an n-cell ${E_1}^n$, and it can be assumed that, in the modification ϕ carrying M_1 into M_2, the complement of ${E_1}^n$ in M_1 is left unchanged. In the proof of this lemma, therefore, nothing is lost if $M_1 - {E_1}^n$ is replaced by a

second n-cell E_2^n so that $E_1^n \cup E_2^n$ is an n-sphere. The modifications are to be carried out on this sphere in such a way that E_2^n is left unchanged, that is to say, so that a neighbourhood of some point is left unchanged.

At this stage the phrase used above, "natural product structure" in a neighbourhood of E^{r+1}, can be explained. The idea is that, when $M_1 - E_1^n$ is replaced by E_2^n to form the sphere $S^n = E_1^n \cup E_2^n$, and then when the neighbourhood B_1 of S^r is removed, the remainder of S^n will be a product $E_1^{r+1} \times S^{n-r-1}$ having the cell E_1^{r+1} as one of its cross-sections. The normal neighbourhood of E^{r+1} will then be the product $E^{r+1} \times U$, where U is a cellular neighbourhood on S^{n-r-1}, with B_1 added on.

The proof of the lemma will now be completed by performing the modifications ϕ and ϕ', related as described above, on the n-sphere S^n, and showing that the final result is again S^n.

S^n can be written as $B_1 \cup (E_1^{r+1} \times S^{n-r-1}) = (S^r \times E^{n-r}) \cup (E_1^{r+1} \times S^{n-r-1})$, where the boundaries of the two products are identified. A point (p, q) is to be selected in the interior of $(E_1^{r+1} \times S^{n-r-1})$, and it is to be checked that at each stage a neighbourhood of (p, q) is left invariant. The modification ϕ replaces B_1 by a product $E_2^{r+1} \times S^{n-r-1}$. Thus, with the boundaries of the products identified, $M_2 = (E_1^{r+1} \times S^{n-r-1}) \cup (E_2^{r+1} \times S^{n-r-1})$. It is clear that a neighbourhood of (p, q) has been left invariant here. Also the identification of the boundaries of the products is such that M_2 is homeomorphic to $S^{r+1} \times S^{n-r-1}$. Now S^{n-r-1} can be written as a union $E_1^{n-r-1} \cup E_2^{n-r-1}$ of two cells, with q in the interior of E_2^{n-r-1}. Thus, the boundaries of the products being identified, $M_2 = (S^{r+1} \times E_1^{n-r-1}) \cup (S^{r+1} \times E_2^{n-r-1})$, and the first product is the normal bundle of S^{r+1}. Thus ϕ' consists in replacing this product by $(E^{r+2} \times S^{n-r-2})$, and the result is S^n; also in the process a neighbourhood of (p, q) is left invariant, and so the proof is completed.

If the situation described in the above lemma holds, the modification ϕ' will be called complementary to ϕ.

One case in which this situation will always hold is where ϕ is a $(0, n-1)$-modification of orientable type. Thus if only the result of a sequence of modifications is of interest (and not the manifold bounded by the initial and final states) every orientable $(0, n-1)$-modification can be replaced by a $(n-2, 1)$-modification.

An important special case of this result is obtained by taking M_1 to be an orientable 3-dimensional manifold. According to Thom's theory of cobounding manifolds, M_1 is the boundary of an orientable 4-dimensional manifold. Hence, by Theorem 2, M_1 can be obtained from a 3-sphere M_2 by a sequence of $(0, 2)$-, $(1, 1)$-, and $(2, 0)$-modifications, those of types $(0, 2)$ and $(2,0)$ all being orientable. By the result just obtained, the modifications of types $(0, 2)$ and $(2, 0)$ can all be replaced by modifications of type $(1, 1)$. Translating into simple geometrical language the meaning of a $(1,1)$-modification, the following theorem is proved, giving an affirmative answer to a problem of Bing **(2)**:

THEOREM 6. *Any orientable 3-manifold can be obtained from a 3-sphere by removing a finite number of disjoint tori and refilling the resulting holes by tori with suitable identification of the boundary surfaces.*

6. Killing the fundamental group. The object of this section is to show that a manifold which is orientable and of dimension n can always be carried into a simply connected manifold by a finite sequence of spherical modifications of type $(1, n-2)$. This having been done, the next section will show how, under certain conditions, this process can be extended to one which will kill all the homotopy, or what in this context is the same thing, the homology groups up to the dimension $n-1$.

The results of this section will be obtained by comparing the fundamental groups of two orientable n-dimensional manifolds M_1 and M_2 which are related by a single spherical modification ϕ of type $(r, n-r-1)$ (necessarily orientable in case $r = 0$ or $n-1$). As in §2, M_1 and M_2 together will constitute the boundary of an $(n+1)$-dimensional manifold M which can be assumed to lie on an $(n+1)$-dimensional real algebraic variety in Euclidean N-space. It is convenient here to arrange the co-ordinates in such a way that M_1 and M_2 are, respectively, the sections of M by the hyperplanes $x_N = -1$ and $x_N = 1$, while $x_N = 0$ is the singular section of M corresponding to the modification leading from M_1 to M_2. The singular point P of this section can be taken as origin. As in § 2 there will be a family F of curves cutting transversally across the sections of M by the hyperplanes $x_N = c$, except at P. The members of the family F passing through P form two cells E^{r+1} and E^{n-r}, the former lying in the set $x_N \leqslant 0$ and having as its boundary the sphere S^r in M_1 shrunk by the modification ϕ, while the latter lies in $x_N \geqslant 0$ and has as its boundary the sphere S^{n-r-1} in M_2 introduced by ϕ.

The most convenient way of comparing the fundamental groups of M_1 and M_2 is to compare them both with that of M_0, the section of M by $x_N = 0$. This will be done by means of the two mappings $f_i : M_i \to M_0$ $(i = 1, 2)$ defined by setting $f_i(p)$ equal to the point on M_0 and on the curve of F through p. These are continuous mappings (**10**, ch. II), and so induce homomorphisms $f_{i*} : \pi_1(M_i) \to \pi_1(M_0)$ $(i = 1, 2)$. Here π_1 denotes the fundamental group, and in the meantime M_1 and M_2 will be assumed to be connected. The following lemma will now be proved.

LEMMA 6.1. (1) *For $1 < r \leqslant n-1$, f_{1*} is an isomorphism onto.*

(2) *For $r = 1$, f_{1*} is onto and its kernel is generated by the image of $\pi_1(S^r)$ in $\pi_1(M_1)$ induced by the inclusion mapping.*

(3) *For $r = 0$, f_{1*} is an isomorphism into.*

Proof. Let α be a closed path on M_1 beginning and ending at a base point p on S^r, and suppose that $f_1(\alpha)$ is homotopic to a constant on M_0 with respect to the fixed base point $P = f_1(p)$. It is clear then that α is homotopic to a

constant on M with respect to the fixed base point p. That is to say, there is a continuous mapping h of a 2-cell E^2 into M such that the restriction of h to the circumference S^1 of E^2 coincides with α (S^1 is being identified with a line segment with ends joined; this is really a description of free homotopy, but for the present purpose no distinction need be made). Now h can be assumed to be an algebraic mapping. This is done by noting that, under the given h, co-ordinates in the ambient N-space are given as continuous functions of the co-ordinates in a 2-space containing E^2. Approximating these functions by polynomials and then projecting normally into M the required result is obtained **(4)**. At the same time h can be adjusted so that $h(E^2)$, now a piece of algebraic surface in M, bears a simple relation to E^{r+1} and E^{n-r}, which can themselves be assumed to be pieces of algebraic subvarieties of M. Namely, it can be assumed that, if $0 < r < n-1$, $h(E^2)$ meets $E^{r+1} \cup E^{n-r}$ in at most finitely many points, while if $r = 0$ or $n-1$ the intersection may also include some arcs of algebraic curves. These cases will now be considered in more detail.

First take the case where $0 < r < n-1$. When the adjustments described above have been made it can be assumed that there is at most a finite set $P_1, P_2, \ldots, P_m$ of points in the interior of E^2 such that $h(P_i)$ is on E^{r+1} or E^{n-r}. If the adjustment to h is sufficiently small the new path α will of course be homotopic to the original one. It can also clearly be arranged that exactly one point q of the boundary S^1 of E^2 is mapped on p by h. Now let U be a small preassigned neighbourhood of P in M, and let W be the point-set union of all the curves of the family F which meet U. It is not hard to see that W is a neighbourhood of $E^{r+1} \cup E^{n-r}$ in M. Since h is continuous it follows that there are neighbourhoods $U_1, U_2, \ldots, U_m$ of $P_1, P_2, \ldots, P_m$ in E^2, which can in fact be assumed to be non-overlapping circular discs, such that for each i, $h(U_i) \subset W$. From q draw an arc β_i to some point on the circumference of U_i, for each i, arranging that the β_i do not meet each other except at q. Let β be the closed path on E^2 starting at q and going along β_i, round the circumference of U_i and back along β_i for each i in turn. This can be done so that β is homotopic on $E^2 - \bigcup U_i$ to the path which makes a single circuit of S^1. It then follows that $\alpha' = h(\beta)$ is a path on M homotopic in $M - E^{r+1} - E^{n-r}$ to α, with respect to the fixed base point p. In fact the deformation of α into α' is carried out in $M - W$, with possibly a small neighbourhood of p added on. But, making use of the family F of curves, it can be seen that $M_1 - (M_1 \cap W)$, along with a small neighbourhood of p, is a deformation retract of this set (cf. **(10)**, p. 17), and from this it follows that α is homotopic in M_1, with respect to the base point p, to the path $g(\alpha')$, where g maps a point t of $M - W$ on the end point, on M_1, of the curve of F through t. $g(\alpha')$ is a product of paths of the type $\gamma_i \alpha_i \gamma_i^{-1}$ where $\gamma_i = gh(\beta_i)$, and the α_i are closed paths in a small neighbourhood of S^r, a neighbourhood which can be assumed to be a product of S^r by a cell. Since $r > 0$, an easy transformation makes the γ_i into closed paths based on p.

Then if $r > 1$, the α_i are all homotopic to a constant on M_1 (in fact in a neighbourhood of S^r), and so in this case it has been shown that the kernel of f_{1*} is the identity. On the other hand, if $r = 1$, the α_i represent elements of the injection image of $\pi_1(S^1)$, as required in the statement of the lemma.

The kernel of f_{1*} must now be shown to be the identity in the cases $r = 0$ and $r = n - 1$. When $r = 0$, $h(E^2)$ can be assumed to meet E^{r+1}, a 1-cell, only at the point p, $h(S^1)$ will not meet E^{n-r}, but $h(E^2)$ may meet E^{n-r} in some curves. In this case, in addition to the points P_i appearing in the above discussion, there may be some algebraic curves in the interior of E^2 carried by h into E^{n-r}. Since h is continuous, it is in this case possible to find a finite number of simple closed loops C_i in E^2, each surrounding one or more of these curves, and each lying within such a small neighbourhood of these curves that $h(C_i) \subset W$ for each i. The P_i not already surrounded by the C_j are to be given neighbourhoods U_i as before, and the U_i and C_j are not to meet each other. The argument as above is then repeated, using the C_i along with the circumferences of the U_j.

The case $r = n - 1$ is a little more complicated. $h(E^2)$ will meet E^{n-r} at most in a finite number of points (and this need only happen if $n = 2$), but it may meet E^{r+1} in both isolated points and in pieces of algebraic curves, some of which may be arcs with end points on α. The inverse images of these arcs will be arcs of algebraic curves with end points on S^1. A preliminary adjustment will be made this time, deforming the mapping h in such a way that all these end points coincide with q. There are now in E^2 isolated points, isolated curves in the interior of E^2, and a set of curves forming a connected set containing q, all mapped into $E^{r+1} \cup E^{n-r}$ by h. The isolated points and curves in the interior of E^2 are to be treated as in the case $r = 0$, and the remaining curve is to be surrounded by a simple closed loop beginning and ending at q and lying in such a small neighbourhood of the curve that it is mapped by h into W. This loop is to be included in the product of paths forming β, and the rest of the argument is the same as before.

To complete the proof of the lemma it must be shown that f_{1*} is onto except in the case $r = 0$; it obviously will fail to be onto in this case. If then $r \neq 0$, let α be a closed path on M_0, and it is convenient this time to take as base point for closed paths a point Q different from P. α is then homotopic in M to a path α_1 not meeting E^{n-r}; this is possible since $r \neq 0$. Let α_2 and α_3 be the projections of α_1 on M_0 and M_1 respectively along the curves of F. The point $f_1^{-1}(Q)$ is well defined and will be taken as base point for closed paths on M_1. Clearly $\alpha_2 = f_1(\alpha_3)$. On the other hand, α_2 is homotopic in M, with respect to the base point Q, to α_1 and hence to α. But, using the curves of F, M_0 is a deformation retract of M (**10**, ch. I, § 4) and so α_2 and α are homotopic in M_0. Hence f_{1*} carries the homotopy class of α_3 in M_1 into that of α in M_0, and this shows f_{1*} to be onto for $r \neq 0$. This completes the proof of the lemma.

The case in which M_1 (or similarly M_2) is not connected is dealt with as follows:

LEMMA 6.2. *Continuing with the notation of the last lemma, let ϕ be a $(0, n-1)$-modification, let M_2 be connected but let M_1 consist of two connected components M_1' and M_1''. Then the fundamental group of M_0 is the free product of the images under f_{1*} of those of M_1' and M_1''.*

Proof. This is a well known result, but is also easy to derive in the manner of the last lemma.

Applying the above lemmas also to f_{2*} and putting the results together, the following theorem is at once obtained.

THEOREM 7. *Let M_1 and M_2 be two n-dimensional orientable differentiable manifolds related by a spherical modification ϕ of type $(r, n-r-1)$. Then*

(1) *if $1 < r < n-2$, $\pi_1(M_1)$ and $\pi_1(M_2)$ are isomorphic under $f_{2*}^{-1}f_{1*}$. (This can only happen if $n > 4$.)*

(2) *If $r = 1$ and $n > 3$, $f_{2*}^{-1}f_{1*}$ is a homomorphism of $\pi_1(M_1)$ onto $\pi_1(M_2)$ with kernel generated by the image of $\pi_1(S^1)$ induced by the inclusion of S^1 in M_1, S^1 being the 1-sphere shrunk by ϕ.*

(3) *If $r = 0$ and $n > 2$, and M_1 is connected, $f_{2*}^{-1}f_{1*}$ is an isomorphism into. If M_1 has two components, $\pi_1(M_2)$ is the free product of their fundamental groups.*

Complementary results to (2) and (3) can of course be obtained by taking $r = n-1$ or $n-2$. The condition $n > 2$ in (3) is no great obstacle, as modifications on a surface are rather a trivial matter. On the other hand the restriction $n > 3$ in (2) shows up one of the essential difficulties of the 3-dimensional case, where a modification which shrinks one circle simply has the effect of introducing another.

Suppose now that M_1 is a compact orientable differentiable manifold of dimension $n > 3$. $\pi_1(M_1)$ is a finitely generated group in this case, and the generators can be assumed to be carried by a finite collection of disjoint 1-spheres differentiably and, of course, directly embedded in M_1. Performing the modifications which shrink these 1-spheres, and using part (2) of the above theorem, we have the following theorem.

THEOREM 8. *An orientable compact differentiable manifold of dimension > 3 can be made simply connected by a finite sequence of $(1, n-2)$-modifications.*

Note that, according to Theorem 6 and the remarks preceding it, the condition $n > 3$ can be dropped. But there is no guarantee in the case of $n = 3$ that the modifications involved correspond to a systematic killing of the generators of the fundamental group.

7. Killing the homology groups. The aim of this section is to give a partial extension of the results of the last section to the homology and homotopy groups of dimension higher than the first. The idea is that a cycle carried by a directly embedded sphere can be annulled by the modification

which shrinks that sphere. But the condition imposed here on the cycle is a rather strong one, and so no sort of complete theory is possible until the situation has been analysed in much greater detail. The ideal result would be to achieve a complete "killing" by adding to the given manifold suitable auxiliary manifolds, namely representatives of the generators of the Thom cobounding groups but, in the meantime, a few of the simpler cases will be treated.

Let ϕ be a spherical modification of type $(r, n-r-1)$ carrying the differentiable manifold M_1 into M_2, shrinking the sphere $S^r \subset M_1$ and introducing $S^{n-r-1} \subset M_2$. Let B_1 and B_2 be the normal bundles of S^r and S^{n-r-1} in M_1 and M_2, both, of course, topological products. Using singular homology with integral coefficients, an application of the homotopy and excision theorems shows that $H_p(M_1, S^r) \cong H_p(M_1 - B_1, \mathrm{Fr}B_1)$, and $H_p(M_2, S^{n-r-1}) \cong H_p(M_2 - B_2, \mathrm{Fr}B_2)$, for all p. On the other hand ϕ induces a homeomorphism between $M_1 - B_1$ and $M_2 - B_2$, and so it follows that $H_p(M_1, S^r) \cong H_p(M_2, S^{n-r-1})$ for all p. The results to be obtained now depend on the examination of the following diagram, in which the horizontal lines are the appropriate homology sequences:

$$
\begin{aligned}
(7) \qquad & \to H_p(S^r) \xrightarrow{i_p} H_p(M_1) \xrightarrow{j_p} H_p(M_1, S^r) \xrightarrow{\partial_p} H_{p-1}(S^r) \to \\
& \qquad\qquad\qquad\qquad\qquad\qquad \| \\
& \to H_p(S^{n-r-1}) \xrightarrow[i'_p]{} H_p(M_2) \xrightarrow[j'_p]{} H_p(M_2, S^{n-r-1}) \xrightarrow[\partial'_p]{} H_{p-1}(S^{n-r-1}) \to .
\end{aligned}
$$

The proofs of the following lemmas are immediate.

LEMMA 7.1. *In the above notation if* $2r < n-1$ *(that is* $r < n-r-1$*) then* $H_p(M_1) \cong H_p(M_2)$ for $p < r$ *and* $H_r(M_2) \cong H_r(M_1)/i_r H_r(S^r)$.

Obviously there is a complementary result for $r > n-r-1$, amounting simply to looking at ϕ as leading from M_2 to M_1. If in the lemma just proved S^r carries a representative of some generator of $H_r(M_1)$, then the lemma shows that the effect of ϕ is to annul that generator.

LEMMA 7.2. *If* $r+1 < p < n-r-1$, $H_p(M_1) \cong H_p(M_2)$, *and except when* n *is even and equal to* $2(r+1)$, $H_{r+1}(M_1)$ *can be identified with a subgroup of* $H_{r+1}(M_2)$ *and the quotient group is isomorphic to the kernel of* i_r.

In particular, this shows that, if the cycle carried by S^r is homologous to zero in M_1 or is a torsion cycle, the effect of the modification (with the exception noted) is to add another generator to the $(r+1)$st homology group. These two lemmas show two of the characteristic ways in which a modification can affect homology. Note that, if M_1 and M_2 are simply connected and $r > 1$ (in addition to the conditions already imposed on it) then the above results, by the Hurewicz isomorphism theorem, can be interpreted in terms of the homotopy groups provided that all the lower dimensional homology groups are already known to be zero.

The cases in which the condition $2r < n - 1$ fails will now be examined. This will have to be done separately in the two cases n odd and n even. First consider an odd value $2m + 1$ of the dimension; the case to be looked at then corresponds to the value m of r.

LEMMA 7.3. *In the situation just described, $H_p(M_2) \cong H_p(M_1)$ for all $p < m$. If the image of i_m is not of finite order in $H_m(M_1)$ then the effect of ϕ on M_1 is to reduce the mth Betti number by 1, but possibly to introduce a new torsion cycle.*

Proof. The first statement, concerning $p < m$ follows at once from the diagram (7). Next, if the image of i_m is not of finite order in $H_m(M_1)$, the fundamental cycle α of S^m will be homologous in M_1 to $k\alpha_1$, where k is an integer and α_1 belongs to a Betti basis for M_1. Using a dual basis, it follows that there is a cycle β on M_1 such that $\beta \,.\, \alpha = k$. Now β can be chosen as a linear combination of singular simplexes on M_1 each of which either does not meet S^m or has exactly one interior point in common with S^m. If the latter simplexes are removed a relative cycle of $M_1 - B_1$ modulo $\mathrm{Fr}B_1$ is obtained whose boundary is easily seen to be $k\gamma$, where γ is a fundamental cycle of the m-sphere S_1^m in M_2 introduced by the modification. Clearly $k\gamma$ is homologous to zero in M_2. Now the diagram (7) gives the isomorphism

$$H_m(M_1)/i_m H_m(S^m) \cong H_m(M_2)/i_m' H_m(S_1^m).$$

Since the images of i_m and i_m' have generators represented respectively by α and γ, the result stated follows at once.

There is obviously a complementary result to the above, starting with the assumption that α is a torsion cycle in M_1; this is not an essentially different result, but simply consists in reversing the parts played by M_1 and M_2 in the above.

Consider next an even value $2m$ for n. The inequality $2r < n - 1$ is equivalent to $r < m$, and so again the case requiring special attention is $r = m$.

LEMMA 7.4. *If the image of i_m is not of finite order the effect of the modification is to decrease the mth Betti number by 2.*

Proof. If α is the fundamental cycle of S^m there is a cycle β on M_1 such that $\beta \,.\, \alpha \neq 0$. Then, reasoning as in the last lemma, it follows that the modification annuls the homology classes, over rational numbers, of α and β; these classes are certainly different, for, since S^m is directly embedded, $\alpha \,.\, \alpha = 0$.

Lemma 7.4 could be formulated more completely by describing the effect of the modification on torsion, but as there is no immediate application it does not seem worth while. In any case the most suitable situation for applying this result would be where the lower dimensional homology groups were all zero, when the m-dimensional torsion group would also automatically vanish.

References

1. A. Aeppli, *Modifikationen von reellen und komplexen Mannigfaltigkeiten*, Comm. Math. Helv., *31* (1957), 219–301.
2. R. H. Bing, *Necessary and sufficient conditions that a 3-manifold be S^3*, Ann. Math., *68* (1958), 17–37.
3. F. Hirzebruch, *Neue topologische Methoden in der algebraischen Geometrie*, Ergebnisse der Mathematik (Berlin, 1956).
4. J. Nash, *Real algebraic manifolds*, Ann. Math., *56* (1952), 405–421.
5. P. Samuel, *Sur l'algébricité de certains points singuliers*, J. math. pures et appl., *35* (1956), 1–6.
6. H. Seifert and W. Threlfall, *Lehrbuch der Topologie* (New York, 1934).
7. ——— *Variationsrechnung im Grossen* (New York, 1951).
8. R. Thom, *Quelques propriétés globales des variétés différentiables*, Comm. Math. Helv., *28* (1954), 17–86.
9. A. H. Wallace, *Algebraic approximation of manifolds*, Proc. Lond. Math. Soc., *7* (1957), 196–210.
10. ——— *Homology theory on algebraic varieties*, (London, 1957).

Indiana University